

# **I AM; I EXIST**

**AN EXPLORATION OF WHAT THE SELF IS AND HOW IT IS  
CONSTITUTED**

Thesis submitted for the degree of Doctor of Philosophy

January, 2014

Yvonne S. Egege  
Philosophy Department  
School of Humanities  
Flinders University of South Australia



'I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.'

Signed \_\_\_\_\_

Yvonne S. Egege, 2014



## ACKNOWLEDGEMENTS

This PhD thesis journey began in 1994 when I first undertook my Masters. I completed my Masters in 2002 on the sense of self. This only served to stimulate my interest in what the self is, a topic that I have never tired of. This is a thesis I wanted to write, even though I ran out of space to say all I think should be said. In the course of that journey I have been helped, encouraged and supported by many people, both personally and professionally.

First, this thesis could not have been written without the support of Flinders University and its staff. They have given endless support and useful advice when needed, providing resources, and allowing me to flexibly manage both work and study. I was able to access a huge range of expert advice on Research Higher Degree matters through the RHD Professional Development Program and other RHD support mechanisms whenever I needed to. Being awarded an APA scholarship in 2008 was invaluable and enabled the bulk of the first draft to be written.

I would like to thank my supervisors, Rodney Allen, Ian Ravenscroft and George Couvalis for all their feedback and intellectual engagement. It is always important to have your ideas respected and to engage in genuine debate. In particular, Ian's intellectual rigour and honest critiquing of my work were invaluable, and I am sure this interaction improved the quality of the final draft. Many of his useful comments are recorded in the text. Special thanks have to go to Rodney for his consistent support and encouragement. He kept me going at times when I doubted the quality of my ideas and my capacity to write. I would also like to thank my academic supervisor Salah Kutieleh and to acknowledge the friendship and consideration he has given me during this process. Without his support I would not have been able to manage my professional work alongside the PhD. I want to also thank my colleagues and my many students for their encouragement and faith in me.

My family have always been very supportive of me writing a PhD. They were particularly considerate in the last few months when tempers were a bit short. Thank you. Special thanks have to go to my daughter Sarah for helping me in the final frustrating stages; editing, formatting and proof-reading on demand. She did a great job. But my greatest thanks have to go to my husband David. He has been a constant

feature in my life, keeping things going while I worked on my thesis, often acting as gatekeeper. He supported me when I didn't know it and even when I didn't acknowledge it. This is to acknowledge the sacrifices he has made and the time he has invested in me in support of my journey.

Y.S. Egege,  
January, 2014

# Table of Contents

<b>THESIS SUMMARY .....</b>	<b>11</b>
<b>CHAPTER 1 METAPHYSICS AND METHODOLOGY .....</b>	<b>13</b>
1.1 Background: The Cartesian legacy and its implications .....	13
1.2 The framework of the debate: the rejection of the Cartesian self and the problem with metaphysics .....	15
1.3 The self and personal identity .....	22
1.4 The narrative self.....	24
1.5 Neuroscience and neuropathologies of the self.....	25
1.6 Physicalist conceptions of the self: Dennett and Metzinger .....	26
1.7 Strawson's selves .....	28
1.8 Damasio, the embodied self and the self in the brain .....	29
1.9 Terminology.....	31
<b>CHAPTER 2 PERSONAL IDENTITY AND THE SELF – FREEING THE SELF FROM ITS IDENTITY .....</b>	<b>33</b>
2.1 Introduction .....	33
2.2 Why the 'self' and 'personal identity' are treated the same.....	35
2.3 The issues of personal identity .....	37
2.4 Locke on personal identity .....	38
2.5 Hume on the non-existence of selves and persons.....	41
2.6 Parfit and impersonal personal identity.....	44
2.7 Critique of Parfit's 'no subject' personal identity.....	47
2.8 Personal identity and Ipseity .....	58
2.9 Psychological continuity and the problem of identity.....	61
2.10 Identity criteria for persons .....	68
2.11 Conclusion.....	70
<b>CHAPTER 3 THE NARRATIVE SELF OR THE SELF-AS-NARRATIVE .....</b>	<b>75</b>
3.1 Introduction: Personal identity and narrative identity.....	75
3.2 Positing the self as a narrative and self-narrating as normative.....	78
3.3 Schechtmann: the self, personal identity and the construction of persons.....	81
3.4 Velleman and multiple selves .....	85
3.5 Dennett and the centre of narrative gravity.....	91
3.6 Critique of Dennett, Schechtmann and Velleman.....	96
3.7 General problems with the self-as-narrative account of the self.....	100
3.8 Narrative, self-identity and authenticity.....	103
3.9 Zahavi, self-phenomenology, and the role of ipseity in selfhood: the limitations of the Narrative self.....	108
3.10 Conclusion – the self is not its narrative .....	113
<b>CHAPTER 4 THE SELF AND ITS PATHOLOGIES .....</b>	<b>117</b>
4.1 Introduction- the issue of unity .....	117
4.2 How pathologies have an impact on the self/ the sense of self.....	119
4.3 Schizophrenia and delusions of control .....	122

4.4 Assessing the credibility of current explanations of the mechanisms of schizophrenia .....132

4.5 Implications for the self and sense of self .....135

4.6 Dissociative Identity Disorder (DID) .....139

4.7 Life as a multiple .....141

4.8 How real are multiples and DID? .....142

4.9 Implications for the singular self .....148

4.10 The case of multiples does not argue against the unity of selves .....152

4.11 Other DID's and their impact on self-world relations .....154

4.12 Cotard's and related derealisation syndromes: their neurological basis .....157

4.13 Implications for the self, self-knowledge and use of the indexical .....162

4.14 The self in the brain .....168

**CHAPTER 5 PHYSICALIST CONCEPTIONS OF THE SELF: DENNETT'S MEME MACHINE AND METZINGER'S PHENOMENAL SELF MODEL .....173**

5.1 Physicalist conceptions of the self and phenomenology .....173

5.2 Dennett's explanation for the existence of the mind, self and consciousness .....175

5.3 Dennett's model of consciousness .....177

5.4 Dennett's two selves .....181

5.5 Problems with Dennett's model of the self .....185

5.6 Why Dennett can't argue against phenomenal consciousness .....191

5.7 Metzinger, phenomenal scepticism and positing the Phenomenal Self Model (PSM) .....194

5.8 How the PSM becomes a 'self' .....198

5.9 Problems within Metzinger's model : Phenomenology and embodiment .....202

5.10 Problem with Metzinger: epistemic scepticism and ontology .....208

5.11 Why can't the PSM be an alternative to the Cartesian self? .....210

5.12 Metzinger and phenomenology .....212

5.13 Physicalism and reduction .....215

**CHAPTER 6 STRAWSON'S SELVES .....223**

6.1 Introduction: an alternative position .....223

6.2 Strawson's position .....224

6.3 Strawson's method .....226

6.4 The SESMET .....230

6.5 Strawson's metaphysics .....233

6.6 Strawson's argument for selves as syseles .....236

6.7 Critiquing Strawson's minimal self: points of agreement .....237

6.8 Strawson's selves: points of disagreement .....239

6.9 The momentary/minimal self is not enough to be a self .....245

6.10 Phenomenology and the experience of the body .....248

6.11 Discontinuous or continuous selves? .....252

**CHAPTER 7 DAMASIO AND THE SELF AS A SYSTEM .....255**

7.1 The importance of the phenomenology of the self .....255

7.2 Damasio and the beginning of the self .....260

7.3 Damasio's self-model and the role of body maps .....266

7.4 Problems with Damasio's self 'process' .....277



7.5	Consequences of Damasio’s model: the self-as-system.....	282
7.6	Establishing the self as object .....	284
7.7	Concluding remarks .....	286
<b>BIBLIOGRAPHY .....</b>		<b>290</b>
<b>APPENDIX .....</b>		<b>308</b>
	Letter from Dennett .....	308



## THESIS SUMMARY

Since Descartes, there has been an ongoing debate about the self, both in terms of what it is and whether or not it can be said to exist. Descartes himself considered the self to be a real existent thing, albeit a non-physical immutable substance. For him, the self, the soul and the mind were the same thing, comprising both the essential, eternal kernel of our identity and the ‘I’ of thought. The problems surrounding the Cartesian self are well-known. While the existence of the mind is widely accepted, although its status as a distinct entity is also debated, positing the self as an existent thing is highly contentious. Very few physicalist theories treat the self as a distinct existing entity in its own right, instead treating it as either identical to personal identity or persons, or considering it as an emergent socio-cultural narrative.

In this thesis, I argue that it would be wrong to treat the self as identical to persons or personal identity or to reduce it to just the having of a self-narrative. I argue that the self can be considered an entity in its own right and that it would be as much an existent thing as, for example, a teaspoon or a leaf. I argue for this in the following way; I claim that the ontological grounds presented for *not* considering the self an existent concrete object can apply equally to any complex object or artefact. Similarly, the self is not unique in its lack of determinate identity conditions or in its sometimes indeterminate persistence conditions. I then set out conditions such that if any thing (such as the self) were to satisfy those conditions, it would have grounds to be considered a concrete entity or object. In the remainder of the thesis, I demonstrate that 1) there is a viable alternative to the Cartesian self and 2) it could satisfy the conditions for objecthood.

In support of my claim, I critique several well-known arguments against the reality of the self (such as Parfit, Schechtmann, Velleman, Dennett, and Metzinger) to point out their respective limitations in dealing with both the phenomenology of the self and/or current neurological findings. In particular, I want to demonstrate that the phenomenology of the self is not fully captured by placing it under the rubric of personal identity or by reducing it to just a self-narrative. I draw on Strawson’s phenomenological framework as support.

I go on to argue that our sense of self is not illusory and that neurological evidence argues against a purely psycho-social or narrative self concept. Based on a discussion of some well-documented neuro-pathologies, I argue that our sense of self is rooted in our physicality rather than our socio-linguistic concepts. This has important implications for A.I. and our understanding of mental health. Using Damasio's model and evidence drawn from the neurosciences, I demonstrate that the sense of self is produced by the concerted actions of various self-identifying and self-informing mechanisms in the body/brain whose concerted actions produce our self-phenomenology. I go on to claim that the complex interrelation of those self-identifying, self-organising and self-directing mechanisms could be enough to treat it as a single entity; this entity could legitimately be called a self, even if minimal. This self-system is primary and fundamental to perception. I conclude that it satisfies the conditions of concrete objecthood stipulated in the early part of the thesis. As such it can be said to exist and be considered as real as any complex concrete entity, such as a toaster or something like the visual system.

# CHAPTER 1

## METAPHYSICS AND METHODOLOGY

Accordingly, having considered it well, and carefully examined everything, I am obliged to conclude and to hold for certain that this proposition, I am, I exist, is necessarily true, every time I pronounce it or conceive it in my mind... But what next is the nature of the self whose existence is so certain?<sup>1</sup>

### 1.1 Background: The Cartesian legacy and its implications

Descartes' dictum 'cogito ergo sum' is probably one of the best known philosophical expressions outside of Academia.<sup>2</sup> His formulation of the Cogito in *The Meditations* and his subsequent analysis of what constitutes the self he claimed to have discovered, established an ongoing philosophical debate which still permeates all contemporary discussions about the mind, consciousness and the existence of the self; the so-called mind/body problem. The effects of his legacy should not be underrated. He effectively split the mind from the body in such a convincing way that the norm in the 21<sup>st</sup> Century is to talk of ourselves as a duality of the physical and the mental.<sup>3</sup> While this doesn't assume a commitment to substance dualism, it does indicate the distinction we draw between the mental realm and our bodies. Not only does the Cartesian self seem to exactly capture our phenomenal experience of being a cognising agent but our laws, social structures and ethics all assume and reflect a similar dualism. It seems it is the mental realm, where the self is considered to sit, that is the defining feature of who one is. In some legal cases, and depending on the circumstances, judgement concerning the extent of one's legal and moral responsibility will be determined not by the co-occurrence of one's body at the site of the disputed action but rather on the condition of one's psychological or mental states at the time.

In many cases, one's guilt or innocence will focus on the extent to which the *person* in question knowingly and willingly committed the act. If it can be demonstrated that the

---

<sup>1</sup> Rene Descartes, *The Philosophical Works of Descartes*, trans. Elizabeth S. Haldane and G. R. T. Ross, Cambridge University Press, Cambridge, 1973, p. 150.

<sup>2</sup> The actual Latin expression is found in the *Principles of Philosophy*, 1644 but not in the *Meditations*, 1641, where he more fully explicates his reasoning for this conclusion.

<sup>3</sup> In this discussion, I am not talking about the views of philosophers or cognitive scientists but of ordinary, everyday 'folk' beliefs about the self, persons and their relationship to our bodies.

person was not their 'normal' mental *self* at the time of the act in question, or that there was a lack of continuity between their prior self and the self who committed the act, then they may not be found culpable. They can escape taking responsibility for the actions their body performed. This is under the presumption that their (mental or psychological) self was either absent or acting out of sync with their 'real' identity. They were not acting in accordance with their normal self.

Such a position demonstrates how, in our current society, our identity or who we really are is considered to be constituted by our *psychological* states more than it is our body. While our everyday identity is generally picked out and tracked by the sameness and continuity of our physical body, our *self* identity is treated as a feature of our mind/self or psychology. These two identities are not always taken to coincide, which means that mind/self can potentially be treated separately from the body. There is an extensive body of literature on the objectification and commodification of the body and its impact on health and identity.<sup>4</sup>

It highlights how our understanding of the mind/body relationship affects our attitude towards and treatment of, for instance, self-psychotic mental illness or disability. Additionally, how the mind, self and body are understood in terms of what (if any) kinds of causal entities they are, along with their inter-connectedness, has important implications for the work being conducted in A.I. and robotics.<sup>5</sup> Such findings will have an impact on the plausibility of building self-conscious automatons. Whether or not we can make self-conscious robots will hinge on how well we understand what it is to be self-conscious entities and what it is that makes us that way.

Establishing whether or not selves exist and what capacities those selves (would/could) have is critical in understanding ourselves both as physically evolving systems and as human beings. The reality or not of selves has implications for the way we interact with other species, the way we organise our societies and the way we develop and deal with

---

<sup>4</sup> See for example Marguerite LaCaze, 'Splitting the Difference: Between Young and Fraser on Identity Politics' in *Feminist Alliances*, Pluto Press, London, 2001.

<sup>5</sup> See John Haugeland, *Artificial Intelligence: the very idea*, MIT Press, Mass, 1987 for an outline of some of the issues involved.

artificial life forms like robots. At the same time, we need to fully understand what we are and what potential capabilities we have to help ground our moral systems, to impart meaning to our ways of life and to design our social structures to reflect these. Part of that capacity will be linked to how autonomous or self-constituting we really are or can be. This has a direct impact on what we consider to be normative behaviour and how we understand and deal with mental and social pathologies. We need to understand to what extent we can act outside of our genetic constraints or our biological blueprints. Understanding the self and its limitations is integral to understanding what is possible for human beings to achieve. If there is no causal self, then we may not have the capacities we think we do. We may not be in control of our actions in the way we think we are. We may not have the capacity to know ourselves in any meaningful way or be able to instigate change in our behaviour. We as a self would not exist. This would have serious consequences for our beliefs in individual moral and social responsibility, and our belief in our capacity to collectively affect human destiny.

## **1.2 The framework of the debate: the rejection of the Cartesian self and the problem with metaphysics**

The mind/body division first presented by Descartes still operates as a conceptual framework in delineating the parameters of the debate about the self. The intuitive strength of Cartesian Dualism is the plausibility of its characterisation of the self. According to Galen Strawson, our phenomenological sense of self universally identifies the self as an ‘inner mental thing.’<sup>6</sup> We neither experience our ‘self’ as a physical thing or object, nor do we consider ourselves as completely identical to our body. We appear to be, in most instances, phenomenological Cartesians.

Despite this seemingly widespread phenomenological concurrence with a Cartesian-like self, positing the existence of a non-physical yet causal entity is problematic.<sup>7</sup> As a consequence, the contemporary debate about the existence (or not) of the self takes the

---

<sup>6</sup> Galen Strawson, ‘The Self’, *Journal of Consciousness Studies*, Vol. 4, nos 5/6, 1997, p. 407.

<sup>7</sup> Problems with Dualism are well-documented and can be found in various philosophy texts on the mind/body problem. See Paul Churchland, Chapter 2, *Matter and Consciousness*, MIT Press, 1988 and John Haugeland, ‘The Saga of the Modern Mind,’ *Artificial Intelligence: The Very Idea*, 1987, pp.33-38 for a brief summary of these, as well as attempts to deal with them.

following as its starting point; it unequivocally rejects the substantive Cartesian self as being both scientifically and philosophically implausible. Whether we experience life from a seemingly Cartesian perspective or not, there are strong arguments to discount the plausibility of such a non-physical self as a causal entity.<sup>8</sup> Likewise, even if we were to reject the non-physical aspect of the Cartesian self, a physically instantiated Cartesian self seems just as unlikely; i.e. some kind of singularity like a physical soul-pearl or a self-neuron in the brain.<sup>9</sup> As Dennett and others have argued, if there is no ‘inner homunculus’ or ‘soul pearl,’ then there is no (Cartesian) self.<sup>10</sup>

The argument seems to go like this. Either there is a (non-physical) Cartesian self or there is a singular physical entity in the brain that acts like a Cartesian self, or there is no *real* self.

- There is no such thing as a Cartesian self.
- There is no singular physical entity in the brain that acts like a Cartesian self.
- Therefore, there is no *real* self.

This is not to say that such writers discount a self of any kind. Many philosophers have put forward alternative epiphenomenal conceptions of the self, such as the socially constructed self or the self-as-narrative.<sup>11</sup> Others claim that the term ‘self’ just refers to the whole human being or is representative of our personal identity or personhood.<sup>12</sup> What the above presents is an argument against the existence of an *indecomposable Cartesian* self, whether physical or non-physical. It assumes that only a Cartesian self

---

<sup>8</sup> While this may appear contradictory, science has frequently shown that what appears to be the case may not be so. The folk may accept the existence of a non-physical self or soul, as some do, but this doesn’t mean that it can be causal within the physical world or that it necessarily exists.

<sup>9</sup> The latter position is close to Strawson’s in *Selves; an essay in revisionary metaphysics*, Clarendon Press, Oxford, 2009. I address his view in detail in Chapter 6 as he posits the existence of a Cartesian-like mental self that is both concrete and object-like. See also David Chalmers, *The Conscious Mind*, OUP, 1996 for a supervenience view of the mental.

<sup>10</sup> Daniel Dennett, *Consciousness Explained*, Little Brown, Boston, 1991. See also Derek Parfit, *Reasons and Persons*, Oxford University Press, Oxford, 1986. I address both views in Chapters 2, 3 and 5.

<sup>11</sup> I discuss these positions in detail in the ensuing chapters but see Dennett, 1991, Parfit, 1986, Marya Schechtman, *The Constitution of Selves*, Ithaca, Cornell University Press, 1996 as exemplars.

<sup>12</sup> Aaron Sloman, ‘*The Self*’ – *a bogus concept*, pp. 1-7, accessed November 2010. <http://www.cs.bham.ac.uk/research/projects/cogaff/misc/the-self.html>. See also David J. Velleman, *Self to Self: selected essays*, Cambridge University Press, New York, 2006; A. Kenny, *The Self*, Marquette, Marquette University Press, 1998.



has the right kind of properties to count as a real, existing self. No other kind of self could be real and/or existent.

I believe this view represents a very narrow ontological position. It seems to assume that any self, for it to exist, must be a singular thing, rather than a complex thing and it must remain that same thing for all time. The consequences of this type of metaphysics can be seen in van Inwagen's ontological claim about everyday items. Other than metaphysical 'simples,' van Inwagen holds the view that only substances that compose a life are real. In discussing composition he says,

[T]he only composite objects are living organisms... [T]herefore, there are no "substances existing by art," such as tables and chairs, and there are no "substances existing by chance," such as stones and bits of stick.<sup>13</sup>

Living creatures are the only genuine composite objects because they cannot be reduced to their parts without explanatory loss, whereas this is not the case with artefacts. This means such things as teaspoons, leaves and selves don't exist.<sup>14</sup> I address this view in more detail in Chapter 5 where I discuss Dennett's position, mentioned above.

Metaphysics should be the ideal framework for establishing the existence or not of selves. It is the role of metaphysics to address questions about what science and ordinary people often assume or take for granted; the reality (or not) of the physical world and the objects within it. Metaphysics is the 'philosophical investigation of the nature, constitution, and structure of reality.'<sup>15</sup> According to Lowe, its purpose is to examine what reality consists of. It is not only about what is logically possible but is about what is; it is supposed to reflect a real state of affairs.<sup>16</sup> The major issue about selves is their status as real existent entities. Do selves exist? If so, what kind of entity are they? Selves

---

<sup>13</sup> Peter van Inwagen, 'Precis of *Material Beings*', *Philosophy and Phenomenological Research*, Vol. 53, no. 3, September 1993, p.2.

<sup>14</sup> In other papers, he includes items produced by plants, such as flowers, fruit and vegetables as being equally non-existent because their loss has no impact on the organism. See P van Inwagen, 'Meta-ontology,' *Erkenntnis* Vol. 48, 1998, pp. 233–250.

<sup>15</sup> P. Butchvarov, 'Metaphysics', *The Cambridge Dictionary of Philosophy*, 2<sup>nd</sup> ed., R. Audi (ed.), Cambridge University Press, U.K., 1999, p.563.

<sup>16</sup> E.J. Lowe, *The Possibility of Metaphysics: Substance, Identity and Time*, Clarendon Press, Oxford, 1998, pp. 1-27.

have been variously categorised as terms of reference, abstractions, processes, properties, social constructions or artefacts, fictional stories and even hallucinations. These categorisations have largely determined the metaphysical status of the self and its status as an existent or non-existent entity.

It is outside the scope of this thesis to go into the metaphysical debate in any detail. However, I want to briefly illustrate that the division of existence into categories of entities is not a simple matter and that, when it comes to what is real, it is not just the self that fails badly. It is far from straightforward what categories there are and which entities fall into which category of existence. In addition, there are very few points of agreement between metaphysicians about what is real.<sup>17</sup> This is clearly illustrated when one looks at the most straightforward everyday category of ‘object.’ It appears that one of the most contentious debates in the literature seems to be in relation to the property/object distinction and their respective status as existing things.

There has been a long debate about the metaphysical status of properties and objects and the relationship they bear to each other, going back to Plato and Aristotle. It seems that both properties and objects can be viewed as either particulars or universals, abstract or concrete. I refer to this debate briefly in Chapter 2. For example, both van Inwagen and Armstrong say that properties-as-universals are real existing abstract entities.<sup>18</sup> In contrast, trope theorists would label properties as particulars and claim they are ‘object-like entities’ that have their own identity-conditions.<sup>19</sup> On their view, objects are just bundles of tropes, hence can be reduced to their properties without loss. There is nothing in addition to properties. This places properties as ontologically primary or more fundamental than objects. In fact, for trope theorists, objects are not considered ontologically real, while the properties are. Heil on the other hand, argues that objects

---

<sup>17</sup> See figure 1 in Lowe, 1998, p.23 for an ontological framework; See Michelene Chi, James Slotta and Nicholas Leeuw, ‘From things to processes: a theory of conceptual change for learning science concepts,’ *Learning and Instruction*, Vol. 4, 1994, pp. 27-43, for an alternative ontological framework that has three basic categories – matter, processes and mental states. All entities fall under one of their sub-categories.

<sup>18</sup> Van Inwagen is a self-professed Platonist. *Metametaphysics: New Essays on the Foundations of Ontology*, edited by David Chalmers, David Manley, and Ryan Wasserman; Armstrong, ‘How do particulars stand to universals?’ in *Oxford Studies in Metaphysics: Vol. 1*, (ed) Dean Zimmerman, OUP, 2004.

<sup>19</sup> Heil, John, *from an Ontological Point of View*, Clarendon Press, Oxford, 2003, p.141.

are not just bundles of properties, and tropes (property particulars) cannot exist independently of the object in which they are manifest. Thus, there is an asymmetric dependence between an object and its properties or modes. For him, the object is primary. For Reichenbach, all objects can be reduced to more fundamental elements, which makes walls and paintings abstract entities.<sup>20</sup>

Variations of these views have some further odd implications. Manley illustrates this issue nicely in the introduction to *Metametaphysics*. In discussing composition and composite objects, he says,

On Cian Dorr's view, composition never takes place. There may be partless particles (simples) arranged in the shape of teacups and turkeys, but there are no teacups or turkeys. On David Lewis's view composition always takes place. So, not only are there teacups and turkeys but teacup-turkeys: spatially scattered objects consisting of one-part dishware and one-part bird. And on Peter van Inwagen's view, simples compose a larger object only when their activity constitutes a life. This gets us turkeys but no teacups. <sup>21</sup>

If things like teacups and turkeys can be said to *not* exist, it is not surprising that something as ineffable as the self has fared so badly in the literature. At the same time, though, given the literature, it is not always clear why the self *in particular* is so soundly rejected as an existing entity. It would seem that if the self were to turn out to be just a bundle of subjective experiences, a bundle of tropes as Hume and Parfit claim, then it is as real or as non-real as any composite everyday object. On the other hand, should the self turn out to be an abstract object (universal), as Dennett proposes, then both van Inwagen and Armstrong would consider it to be real but Reichenbach would not. On the other hand, should the self be considered a property itself (say of a person), as Rudder-Baker argues, then it would be as real as a particular or a trope.<sup>22</sup> For some, this would make it more real than the person (trope bundle) in which it inheres. Then again, should

---

<sup>20</sup> I address his view in Chapter 5. Hans Reichenbach, *Experience and Prediction*, Phoenix Edition, University of Chicago Press, 1961.

<sup>21</sup> David Manley, 'Introduction,' *Metametaphysics: New Essays on the Foundations of Ontology*, (eds) David Chalmers, David Manley, and Ryan Wasserman, OUP, New York, 2009, p.2.

<sup>22</sup> Lynne Rudder Baker, 'Beyond the Cartesian Self,' from a paper presented at *The Cartesian 'Myth of the Ego' and the Analytic/Continental Divide conference*", Radboud University, Nijmegen, 3-4 September, 2010.

the self turn out to be a process rather than a thing, a claim supported by some, then Metzinger for one would argue that this means it is not real.<sup>23</sup>

Rescher, on the other hand, argues that processes precede objects and are, thus, the more fundamental category.<sup>24</sup> In fact, he claims that all objects can be considered processes first and objects second. So it would seem that the self could be considered either real or non-real, depending on one's ontological position in relation to properties, objects and processes. It could be considered as real as a van Inwagen turkey or as non-real as a van Inwagen teacup. Under these categories, a range of everyday items that we think are perfectly real would also have no metaphysical reality.

Scientific entities fair little better. According to Smart, one cannot be committed to the existence of postulated scientific entities without being committed to their ontological status.<sup>25</sup> If an entity is considered real within the framework of a theory, then it is considered ontologically real as well. Yet many of the entities posited in scientific theories are unobservables, like the supposed self, which means there are no direct means of perceiving them, hence they are not easily verifiable. This means that scientific theories rely heavily on inference and induction. Scientific realists argue from the success of their theories to the truth of the entities postulated.<sup>26</sup> Sceptics claim that while past theories may have been successful, some have since been shown to be false.<sup>27</sup> With some past theories, the entities posited have been shown not to exist. Thus, the existence

---

<sup>23</sup> See U. Neisser, 'Five kinds of self knowledge', *Philosophical Psychology*, Vol 1. Issue 1, 1988, pp 35-59. Metzinger says this means it is not an object. *Being No-one: The self model theory of subjectivity*, MIT Press, Mass, 2003.

<sup>24</sup> See Nicholas Rescher, *Process Philosophy: a survey of basic issues*, University of Pittsburgh, Digital Research Library, 2009, <http://digital.library.pitt.edu/cgi-bin/t/text/text-idx>.

<sup>25</sup> J. J. C. Smart, 'Phenomenalism about Sub-Microscopic Objects,' *Philosophy and Scientific Realism*, Routledge and Kegan Paul, London, 1963, p. 33-49.

<sup>26</sup> See G. Couvalis, *The Philosophy of Science: Science and Objectivity*, London, Sage Publications, 1997.

<sup>27</sup> Variations on this argument can be found in several sources. I have drawn on Alan Chalmers, Heil and Chakravartty.

of scientific entities can be doubted as well. This has led to scepticism about the capacity of science to tell us about the world in which we live.<sup>28</sup>

What this discussion suggests is that metaphysical existence is not uncontroversial for all objects; scientific objects, complex objects, artefacts and changing entities such as organisms. The self is not alone in this regard. The same is true for establishing determinate identity conditions over time and change. According to Lowe, there are two inter-related concepts in metaphysics that are fundamental - those of 'object' and 'identity.'<sup>29</sup> If an object can be differentiated from other objects then it has an identifiable identity. Thus an object is any entity that has determinate identity conditions and objects are considered objects because they have determinate identity conditions. Again, as I will demonstrate in the following chapters, establishing identity conditions is not just problematic for the self.

Having said that, if I want to argue that the self may be considered an existent thing like other existent things, I will need to demonstrate that it does exist in some way as an entity-like thing. I will need to set out what conditions it would have to satisfy to be considered an entity. For this, I draw on Lowe's criteria of what constitutes an object such as identity and persistence conditions.<sup>30</sup> According to Lowe, it is a wide-spread metaphysical practice to give some kind of ontological priority to concrete objects. The term 'object' carries assumptions taken from our everyday world of macroscopic objects. We assume they have properties like mass, weight, dimension or even visibility; they are available to the senses, hence real or existent. If the self could plausibly be classified as a concrete object, then it would, at a minimum, be said to have some kind of existence. In Chapter 2, I set out a framework for delineating the conditions that some thing would have to satisfy to be 1) a self and 2) an object, drawing on what Chalmers

---

<sup>28</sup> According to Alan Chalmers, this leads to a form of global anti-realism; we could never know it because we don't directly interface with that reality. Alan Chalmers, *What is this thing called science?* 3<sup>rd</sup> edition, UQP, St. Lucia, 2006, pp. 228-32.

<sup>29</sup> Lowe, 1998, p. 30; Heil, 2003, pp. 180-182. see van Inwagen's 'Precis of Material Beings' 1993.

<sup>30</sup> See E. J. Lowe *The Possibility of Metaphysics: Substance, Identity and Time*, 1998 and *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*, Clarendon Press, Oxford, 2006.

refers to as ‘ordinary existence assertions’ emerging from a commonsense ontology.<sup>31</sup> I return to these in the final chapter.

In the following sections, I outline the chapters in this thesis. I critique a series of standard and not-so-standard accounts about what the self is to demonstrate the authors’ ontological commitments about reality, its impact on how the self is conceived, and the shortcomings of their accounts to adequately address the self-phenomenology that is at the heart of the problem of the self. In the process, I want to demonstrate that, as human beings, we have no choice but to develop selves and to be self-conscious. This is the way we are designed. It is my claim that our selves are rooted in our biology and have emerged as a consequence of our evolutionary history. As such, they are not just a product of social processes nor are they just a narrative. I conclude by adverting to the existence of mechanisms in the brain that contribute to the creation of our self-phenomenology. It is my contention that this complex network could be classified as an object, hence a self. I claim this non-Cartesian self is a viable alternative to the traditional Cartesian self.

### **1.3 The self and personal identity**

As previously stated, most modern conceptions of the self have focused on the problematic issue of what constitutes personal identity. The problem of identity is summarised as follows: If there is such a thing as the self and it is always the same self of which one is conscious from moment to moment, then there should be something about that self that can be identified and re-identified as that same self despite changing circumstances and the passage of time. These are the so-called synchronic and diachronic aspects of identity; what constitutes one’s identity at any given time and the subsequent identity conditions that need to be fulfilled in order for sameness/continuity of identity to be established.

---

<sup>31</sup> Chalmers claims that there is a distinction between ontological theory and commonsense ontology. This will lead to a distinction between ontological existence assertions and ordinary existence assertions which are underpinned by a different set of truth conditions. Mereological sums like turkey-teacups would not exist in a commonsense ontology but turkeys and teacups would. David Chalmers, , ‘Ontological anti-realism,’ *Metametaphysics: New Essays on the Foundations of Ontology*, (eds) David Chalmers, David Manley, and Ryan Wasserman, OUP, New York, 2009, p.12.

According to most philosophers in the field, one cannot establish a continuous same self because one cannot establish determinate identity criteria. There are some circumstances where there just is no fact of the matter about whether  $x$  at  $t_1$  is the same as  $x$  at  $t_2$  or whether it is  $y$  or  $z$  that is the true continuant of  $x$ . It is this difficulty of providing determinate identity criteria for the self over time that has led to a general rejection of the substantive self.<sup>32</sup> As a consequence, the literature very rarely talks about the self. Instead, it focuses on trying to establish robust criteria for personal identity, generally conceived as consisting of some form of psychological criteria plus memory (re Locke) or, following Parfit's formulation, defined in terms of persistent psychological relations between occurrent mental states.<sup>33</sup>

As I mentioned above, establishing determinate identity conditions is problematic for all objects yet this does not prevent us from happily identifying and re-identifying objects with a great deal of accuracy. While it is always possible to be mistaken, this is no more a problem for the self than it is for all objects. I will argue that the issue of personal identity is not critical to the existence or not of the self. I want to separate issues of qualitative (personal) identity from issues of quantitative identity. Qualitative identity is about what *kind* of self one is; whether we are a self or not is irrespective of its particular identity. I may not know *who* I am but I know *that* I am. Thus the self can, conceptually, be separated from its qualitative identity. The experience of selfhood is based in subjectivity not identity. However, quantitative or numerical identity comes into play when we want to establish that the self is the same self, continuous through time and change. Again, it will be my contention that even if there are hypothetical situations where this may not be determinate, this does not in itself negate the existence of the self. That kind of indeterminacy will be true for all entities.

---

<sup>32</sup> By substantive self, I mean a self as a distinguishable substance which independently provides sameness and continuity by its perseverance through time. This is Parfit's 'further fact'. Shoemaker argues that it is akin to a truism to claim that the self is a substance or entity. He sees 'subject' and 'substance' as synonymous. See Sydney Shoemaker, 'Are Selves Substances?' *Self-Knowledge and Self-Identity*, Cornell University Press, 1963. Armstrong claims that Hume disagreed with Locke because of his positing of substrata, a 'thin' particular in which properties inhere. Locke argued that the property of 'extension' made no sense without substance. See David Armstrong, *Universals: an opinionated introduction*, Westview Press, Colorado, 1989.

<sup>33</sup> Derek Parfit, *Reasons and Persons*, Oxford University Press, Oxford, 1986.

## 1.4 The narrative self

According to Rorty, positing the ‘self’ as a narrative or social product can be seen as a natural extension of the empiricist account of personal identity.<sup>34</sup> Introducing a narrative which constructs the self within its social context can go some way to providing the human system with a sense of unity, continuity and identity, albeit artificial. It captures the reality of changing identity and satisfies the desire for self-determination rather than biological determinism. It fits with sociological models which claim that social processes determine the nature of the self, its identity, its traits or even its anti-social tendencies. It also ties in nicely with Eastern beliefs about the no-self.<sup>35</sup>

While most of us don’t consciously construct a story, we do develop a sense of our identity over time which may or may not be entirely accurate. And we can respond with some sort of story when asked to explain our actions. In this sense, our identity can be seen as indeterminate and socially constructed, a sort of narrative constructed from biographical details, behavioural evidence and our personal experience. This may be why the concept of the self as narrative has become an extremely popular and widely accepted explanation of what the self *is*. It also fits nicely with the idea that there is no inner ‘given’ self.<sup>36</sup>

It is my claim in this chapter that the self is not just a narrative and that positing the self as narrative does not do the work it is claimed to do. I outline what I believe to be the short-comings of narrativist accounts and what it is about the self that their respective positions do not satisfactorily explain. Although the narrative can be seen as a tool for providing coherence and understanding to our actions, I argue that our concept of the self is based on our experience of ourselves as a distinct thinking thing, regardless of what its narrative is. Thus, the narrative does not capture why we think and act from a

---

<sup>34</sup> See Amelie Rorty, (ed.) *The Identities of Persons*, University of California Press, California. 1976. For examples see Bronwyn Davies and Rom Harre, ‘Positioning: The Discursive Production of Selves,’

<sup>35</sup> Parfit refers to Buddhist beliefs about the self to support his claim that personal identity is not what matters and that ‘false’ beliefs in the self as a distinct or separate entity lead to excessive self-interest. See Parfit, *Reasons and Persons*, Oxford University Press, Oxford, 1986, chapter 14 and appendices.

<sup>36</sup> See Paul Ricoeur, *Oneself as Another*, trans. Kathleen Blamey, University of Chicago Press, Chicago, 1992, Velleman, 1999 and Marya Schechtman, *Constitution of Selves*, manuscript draft, 2000.



position of 'selfness' in the first place. My critique of the personal identity debate applies equally to narrativist accounts of the self.

In addition, I argue that the concept of the self as narrative is flawed. It does not explain our phenomenology, it does not explain agency and it does not explain rational action. It is used so widely that it is in danger of becoming an empty truism. I conclude that, while the narrative might be a useful analogy, it does little to explain how we come to be self-conscious agents. The self is more than just a story.

## **1.5 Neuroscience and neuropathologies of the self**

The so-called problem of the self, trying to explain or explain away the actual existence of a self, emerges from our self-phenomenology. We claim to have a sense of self. We experience ourselves from a position of selfness and that selfness seems to be distinct from other experiences we have. This self-phenomenology is not identical to one's personal identity and is distinct from one's identity conditions. If the sense of self and identity can be pulled apart, as I suggest, then one could imagine that one could retain a sense of one's self even if one's identity was disparate, unfamiliar or apparently non-continuous. This appears to be the case with amnesiacs and DID sufferers, if their self-reporting is correct. They still retain a sense of self and refer to themselves using the first person indexical but appear to have lapses of consciousness during which they manifest an alternate identity. Conversely, one could imagine cases where one lost that sense of self and yet still retained knowledge of one's identity. Cotard delusion and anosognosia appear to represent this category. They sustain damage to their sense of selfness in quite profound ways yet without any loss of narrative history.<sup>37</sup>

The apparent destruction of or damage to the self and its phenomenology are often taken as evidence that the self is not as intrinsically singular or unified as a Cartesian self should be. The conclusion drawn is that selves don't exist. In this chapter, I discuss

---

<sup>37</sup> Damasio claims that patients with anosognosia retain all their personal biographical details, their 'self narrative' if you will, yet cannot use that information to make decisions about the sorts of things that matter to them. They don't know how they feel or what they want. They are largely indifferent to themselves. Antonio R. Damasio, , Chapter 7, 'Emotions and Feelings,' *Descartes' Error: Emotion, Reason and the Human Brain*, Papermac, London, 1996, p155.

various well-known neuropathologies of the self such as DID, schizophrenia, Cotard's and other paraphrenias. I argue that neuropathologies of the self may not indicate a lack of self or a lack of an intrinsic singular or unified self. Just because something can be damaged does not indicate that the object never existed. Likewise, just because the self can be pulled apart, as in some cases of schizophrenia, does not automatically indicate it was never unified. Thus, they are not in themselves evidence of a non-existent self.

More importantly, these disorders are not purely psychodynamic but have a very specific biological or neurological cause. They offer evidence that the sense of self is strongly linked to, or even generated by, very basic biological mechanisms. It could be argued that this biological or bodily-generated self is responsible for producing very distinctive phenomenal (and cognitive) effects. The findings seem to indicate that primitive biological mechanisms play a critical role in defining ourselves, in creating our sense of self and in meaningful decision-making. Bermudez' work on self-referential mechanisms, Damasio's research on the limbic system, and Young's research on delusions of misidentification offer evidence of a plausible explanation for the persistent phenomenology of the self. If this phenomenology is grounded in bodily or primitive neuro-chemical mechanisms, then one could argue that such mechanisms could be said to constitute a physical instantiation of a self.<sup>38</sup> I conclude that not only does this offer a better explanation of our self and our relation to the world but it argues against a purely narrative conception of the self. It would appear that our self and its phenomenology are not just the product of a story constructed of socio-cultural influences and selected biographical details.

## **1.6 Physicalist conceptions of the self: Dennett and Metzinger**

Kierkegaard believed that investigating the self was outside the realm of the sciences.<sup>39</sup> He believed that there was an explanatory gap between what science could say about the

---

<sup>38</sup> Damasio, 1996; Jose Luis Bermudez, *The Paradox of Self-consciousness*, MIT Press, Bradford, 1998; Andrew Young, *Face and Mind*, OUP, Oxford, 1998.

<sup>39</sup>Søren Kierkegaard, *The Concept of Anxiety* trans. Reidar Thomte, Princeton, NJ, Princeton University Press, 1980.

self and what the experience of subjectivity was like. Science did not seem to capture what it was like to be a self-conscious entity.

In recent years, there have been various attempts to describe or define the self and explain its phenomenology from a non-Cartesian and hence, physicalist perspective, from within science. In this chapter, I address a now (in)famous account of the self presented by Dennett in which he explicated the notion of the self as the Centre of Narrative Gravity.<sup>40</sup> I follow this account with a more recent version presented by Metzinger. I use Metzinger because, like Dennett, he both incorporates the phenomenology then dismisses it as illusory. He replaces the self with his own phenomenal self model (PSM) using a typical computational information-processing model of the brain, mind and self.<sup>41</sup> As such, the two are a good illustration of a particular body of research within philosophy of mind.

Both accounts exemplify a sceptical approach to the self, in which the sense of self is described as an illusion created by an elaborate representational network. Both rest on a dismissal of phenomenal consciousness and, therefore, our phenomenological reports about what it is like to be a self-conscious entity. I argue that neither account adequately addresses the phenomenology nor provides an explanation for the nature of its presence. The respective systems they put forward could operate equally well whether conscious or non-conscious. At the same time, just because the phenomenology is created or caused by some physical process does not make it illusory. Such dismissals of phenomenological reporting can lead to epistemic scepticism about all things, not just the self.

While Metzinger's PSM could count as a self, he argues that it is a *model* of a self-*process* and this is not a real self, nor is it a model of a real self. At the same time, he claims that the PSM is a real entity, discoverable by science. This claim demonstrates a confused ontological position as referred to earlier in this introduction. I go on to claim that there is neurological evidence that the phenomenology itself is not an inference to

---

<sup>40</sup> Dennett, 1991.

<sup>41</sup> Metzinger, 2003.

the best explanation but an actual product of certain specific neurological mechanisms in the body and brain such as the limbic system or parts of the hippocampus. One could argue that it is a matter of opinion whether or not we claim that this phenomenology is an illusion or just how we are as perceiving creatures.

However, if there really is an object in front of me or I really have a damaged foot, then it is not clear in what sense the perceptions I have of these things are illusory. I conclude by claiming that there are better explanations of the self and its phenomenology than those presented by Dennett and Metzinger.

## 1.7 Strawson's selves

As a contrast to Dennett and Metzinger, I closely examine an alternative account of the self put forward by Strawson.<sup>42</sup> Strawson's account is also physicalist. However, he puts forward a model of a Cartesian-like self that he claims could count as a real existent concrete object. He starts by assuming the reality of our self-phenomenology and uses this to ground his definition of a self. Except for its non-physical nature, the self he describes is based on traditional Cartesian self properties. Through a process of elimination, he reduces these features to what he claims are the necessary and sufficient conditions for something to count as a minimal sense of self. In this way he eliminates the problematic issue of diachronic identity.

While Strawson does manage to construct a feasible realist account of the self, he does so by eliminating many of the features that would seem to be part of any viable self, such as agency and temporal continuity. His singular 'self-neuron' is so deplete of defining features that it doesn't seem to warrant the label 'self.' I argue that a singular neuronal cluster is unlikely to have the capacity to be self conscious in the way Strawson needs. I also argue that objecthood requires some form of temporal continuity, at least potentially. Strawson's minimal self, his SESMET, lasts under a second and has no potential to persist in time.

---

<sup>42</sup> Galen Strawson, *Selves: an essay in revisionary metaphysics*, Clarendon Press, Oxford, 2009.

Strawson's account is very much embedded in metaphysics, an approach similar to my own. He is also insistent that any metaphysical claims about the reality of the self should use our sense of self as a limiting constraint. However, while he claims to ground his account in universal phenomenology, he relies on his own less common phenomenology to justify his proposal. He claims to have no personal sense of a temporal self, nor any personal sense of embodiment. I argue that this is not universal and he presumes what he sets out to prove. I conclude that Strawson's self is unviable as a self and that it ends up out of sync with the phenomenology. I claim he underrates the embodied nature of the self and that this is what provides the sense of temporal continuity.

## **1.8 Damasio, the embodied self and the self in the brain**

In this final chapter, I draw on Damasio's embodied account of the self as an exemplar of an alternative to a Cartesian self that appears to accord much more closely with our real-life experience of being self-conscious entities.<sup>43</sup> Damasio, like Strawson, claims that the self is real; it exists. However, his self is based in bodily processes rather than mental ones. He argues that the self is primarily a sensor rather than a thinker and describes how the self comes into being in a series of stages from proto-self through core self to autobiographical self.

Damasio's view of the self is useful in explicating the difference between an embodied, physical self and the mental selves presented by other authors. Unlike other accounts, it is an alternative to the Cartesian self that is not just epiphenomenal. I claim that Damasio puts forward a convincing case for a primitive self as a necessary feature of conscious human life. For Damasio, if there is no self there is no consciousness. Consciousness and selfhood are basic features of human perception and are produced by its unique anatomy/neurology.

I draw on other neurological evidence to demonstrate that there are mechanisms in the brain and body that appear to work together to produce our sense of self. I claim they act as a unified self and that this is how the self is constituted. It is my contention that

---

<sup>43</sup> see Antonio Damasio, *Self Comes to Mind*, 2010 and *Descartes' Error* 1996.

empirical data will eventually demonstrate that this constitutes a complex network of closely interrelated parts such that they act in concert to provide consistency and perceptual coherence. This creates the phenomenology of a singular unified entity which is at the root of our sense of self, our self interest and our self consciousness.

I claim that this complex satisfies the following necessary and sufficient conditions to count as a concrete object:

- Unified boundedness or singularity – it must be identifiable as an entity in itself such that, if complex, it has an inherent unity of parts and that its boundaries can be delineated, enabling differentiation from other entities. It can individuate itself.
- Subjectivity – there must be a subject of experience such that the experiences are not just owned or identified as belonging to x, but they are felt directly by x and as occurring within the parameters of x (selfness and consciousness).
- Synchronic and diachronic identity – for something to count as an entity it must satisfy certain identity conditions. Existence, sameness and persistence over time appear to be essential for this.
- Agency – this relates to an entity's causal capacities. For something to exist, it would have causal powers. For the self to exist, it should play a role in the cognitive network such that it has an impact on actions and is aware of its impact on actions.

If the self were represented by a set of clearly defined and well-integrated mechanisms in the brain, then there seem to be plausible grounds for claiming that such a system is both concrete and object-like. I would want to go further and claim that it would represent a physically-instantiated self. As Strawson is at pains to point out, if there is something that has the properties of a self or which is responsible for the sense of self, then that thing would count as a self. It is my contention that this is the case. Ultimately, it will depend on empirical evidence as to whether there is enough unity or integration between the various self-mechanisms for it to count as a singular thing. I would contend that there seems to be sufficient evidence at this stage to indicate that this might be so and I would predict that the more we find out, the more self-like these systems will appear.

## **1.9 Terminology**

It is important at this point to clarify what I am referring to when I use the term ‘self’ in this thesis, given that the nature of that self has yet to be decided. When I use the term ‘self’ in a general way, as in ‘is there such a thing as a self?’ I am not referring to the traditional Cartesian self. I take it as a given that the Cartesian self does not exist, at least not in the way formulated by Descartes. What I am referring to is the not yet delimited self that is the supposed subject of our self-phenomenology. It refers to that thing that may or may not be responsible for that phenomenology or which may or may not be responsible for our decisions, whatever could potentially do the work of the Cartesian self. As such, it does assume many of the traditional Cartesian properties, in particular that it is some identifiable or distinct thing; it has some kind of unity or singularity such that it can be differentiated from other things, that it can be causal. But at this point, I make no assumptions about its metaphysical status or its objecthood.





## CHAPTER 2

# PERSONAL IDENTITY AND THE SELF – FREEING THE SELF FROM ITS IDENTITY

[O]ne's states of consciousness, one's thoughts and sensations, are ascribed to the very same thing to which these physical characteristics, this physical situation, is ascribed... But they do not explain why I should have the concept of myself at all, why I should ascribe my thoughts and experiences to anything.<sup>1</sup>

### 2.1 Introduction

In the previous chapter, I mentioned the issue of unique (quantitative) identity and continuity of identity over time and how both are problematic for any object, particularly for objects that change substantially over time. One of the main arguments against the existence of selves, any kind of self, has been its supposed lack of determinate identity. It has been extensively argued that there is nothing that appears to constitute 1) an independent self and 2) that same self over time and change. This has led to the claim that there is no self, or that talk of the 'self' is synonymous with talk of one's 'person.' Similarly, the term 'self' is often treated as synonymous with the term 'personal identity.' The posited existence of the self hinges on there being some unique or singularly-identifiable thing that can be said to be a self, distinct from other things (such as the human body or brain). For this reason, it is important to see whether or not some set of identity conditions can be satisfied for any self such that it can be said to exist. This constitutes its numerical identity, that there is a distinct something. An object must have a determinate identity and there must be conditions of its identity that it has to satisfy, in order for it to be considered an object of a particular type and for it to be considered that specific object rather than some other object.

In this chapter, I want to discuss what the identity conditions for selves might be. There is, however, another type of identity that arises when discussing human beings, that of personal identity. Personal identity has been defined as that identity that makes you a particular kind of person as opposed to just a human body or being. Following from Locke, it acknowledges that the identity conditions for humans are not the same as

---

<sup>1</sup> Peter Strawson, *Individuals*, London, Methuen, 1959, pp.331-332.

identity conditions for other objects. It acknowledges the existence of psychological or mental properties. As per the elder Strawson, quoted above, this is what constitutes persons.<sup>2</sup>

In the following sections, I will be examining the extensive and often discordant body of literature on personal identity as it relates to the self. Within this body of literature, it is common practice to subsume the self under the rubric of person or personal identity. Consequently, the term ‘self’ is often referred to, or is used interchangeably with, the term ‘person’ or the term ‘subject,’ both of which become interchangeable with the term ‘I’. Like the elder Strawson, many philosophers accept the existence of persons but argue that the self and the person are one and the same. They argue that there is no *additional* subject of experience that qualifies as the Cartesian (or other) self; we just have persons. I argue that what we mean by ‘self’ and what we mean by ‘person’ are not the same thing; the terms refer to different sets of properties. Whether or not either or both can be considered distinct entities is a different matter and needs to be argued for (or against). It is my contention that the self is more than just its personal or qualitative identity, which means the term ‘personal identity’ is not sufficient to capture what we mean by the term ‘self.’ In this chapter, I will argue that while the self, to be a self, will need to be identified as a *self* rather than something else, what kind of self it is is not relevant. I want to demonstrate that the self can, at least conceptually, be separated from its personal identity and still be a self. If this is the case, then many of the problems raised within the personal identity debate will not be relevant to the identity of the self. It is also my contention that reducing the self to its personal identity ignores the very phenomenology that leads us to posit a self. Additionally, I argue that the sceptical conclusions drawn from the apparent *indeterminacy* of self-identity can apply equally to all of our everyday objects, at least some of the time.

---

<sup>2</sup> I refer to (and have already referred to) Galen Strawson repeatedly throughout this thesis. To avoid confusion, I refer to his father as the elder Strawson.

## 2.2 Why the 'self' and 'personal identity' are treated the same

According to Ricoeur, the focus on identity and personal identity in relation to what the self is (or is not) is a direct result of the way Descartes phrased the Cogito.<sup>3</sup> Ricoeur begins his hermeneutics of the Self by critiquing traditional accounts of the self, starting with the Cogito. According to Ricoeur, there is no direct path to the self. The self is not immediately available to introspection, but is only discovered by a circuitous route of investigation. Pace Descartes, he claims that the 'I' that begins the search, the 'I' that doubts, is not the same thing as the 'I' that Descartes finds, what he calls the Cogito. According to Ricoeur, there is a shift in Descartes from who is the 'I' who thinks, to the actual nature of the 'I', which is about 'what'. The 'who' initially refers to the individual specific singular mental/social identity. The 'what' seems to refer to the type of thing that it is, the sortal term one falls under, along with all its physical attributes. This means that the Cogito moves from the 'who' of the *experiential* being to become the 'what' of description, 'that which doubts, thinks, wills, remembers.'

Ricoeur claims that this has two consequences. First, it transforms the initial personal identity of the 'who' from the very specific to the generalised; the Cogito thus becomes a-temporal, a-historical, impersonal, a free-floating subjectivity rather than a specific who that is (always) embodied and embedded in a very specific historical moment. Secondly, this shift leads Descartes into positing the identity of the Cogito as sameness. According to Ricoeur, Descartes moves from the 'I am' as in the absolutist sense of 'I exist' to the 'I am' of description as in 'I am something'. Thus what I am becomes I am *that* which thinks, doubts, wills, remembers, which is actually I am *the same thing*, the permanent continuous thing which is common to all these diverse actions. Ricoeur claims that this shift has laid the groundwork for all future debate about the nature of the self, including its emphasis on finding 'the same continuous thing' common to all one's experiences, the continuing 'I' of personal identity. Ricoeur claims that this shift, which he feels is erroneous, has inevitably led to first Hume's, then Nietzsche's scepticism about the self.

---

<sup>3</sup> Ricoeur, 1992.

Whether one accepts Ricoeur's critique of the Cogito or not, he is correct in regards to the ensuing focus on identity and the subsequent scepticism this entailed. Philosophers from Locke onward have focussed on the trenchant issue of identity, striving to show in what personal identity consists and what, if anything, remains the same throughout time and change. The lack of success in finding an 'indecomposable simplicity' or determinate identity criteria has led to general scepticism about the actual existence of a substantive Cartesian-like self.<sup>4</sup> It has even led to scepticism about the very existence of persons. This has led to a more general scepticism about the ontological reality of any kind of self.<sup>5</sup>

This scepticism does seem inevitable if one moves away from Descartes' initial formulation of existence (I am) to one of nature. As Ricoeur demonstrates, this is a shift from essence to identity. But if the self does not consist in the personalised 'I' of 'I who am Sandra' or the objective 'I' of 'I that thinks, acts, feels', then in what does it consist and where are we going to find it? Ricoeur claims that the quest for the self will, of necessity, require us to identify a particular self from amongst others, as well as the capacity to re-identify that same self over time. However, the hallmark of self-consciousness is the capacity to be aware of ourselves *as ourselves*, to say 'I am' without need of other identifying support. This dichotomy within the concept of the self is what Ricoeur terms the 'ipse/idem' distinction. He argues that the debate about personal identity represents a focus on 'idem' identity only, which is identity through sameness. Ipse-identity or selfhood is left unexplored.<sup>6</sup>

In the following sections, I want to demonstrate that the self is not reducible to just its idem identity and that it can, in principle, be distinguished from it. This is because there are other fundamental properties of the self, such as Ricoeur's ipseity, (variously referred to as subjectivity, beingness or first-person givenness) which constitute a significant portion of our self phenomenology. As a consequence, I will argue that the

---

<sup>4</sup> As mentioned in Chapter 1, the lack of success in the search for a singularity (as an essential feature of the Cartesian self) has led to scepticism about the self, rather than a reconceptualising of the self.

<sup>5</sup> I have already mentioned Dennett and Metzinger as exemplars of this position.

<sup>6</sup> Ipseity is the experience of oneself as existing or being, the awareness of being a living thing. For references see Ricoeur, 1992 or Gallagher and Zahavi in *Models of the Self*, Imprint Academic, UK, 1999.

problem of finding determinate diachronic identity conditions neither proves nor disproves the existence of a distinct self. Consequently, the problems raised in this section of the debate are largely irrelevant to my argument for an existing self. Where the issue of determinate identity is relevant, it is no more nor less problematic than it is for all objects.<sup>7</sup>

### 2.3 The issues of personal identity

There are two main issues that arise in the debate on personal identity which are important here. Both are a consequence of the Cartesian claim that the self is a permanent, immutable ‘substance’ that is continuous through time. The first conclusion is that the self is nothing over and above its identity; there is no self substance *in addition to* its identity. The second is that there are no determinate identity conditions for one’s personal identity. Both these conclusions come from a position of scepticism towards the existence of the self as 1) a distinct entity and 2) as a same self over time and through change. As someone like Dennett claims, the self is not ‘permanent or unified but disparate and changeable over time.’<sup>8</sup> Dennett says that the self changes inasmuch as its identity changes over time. So the self and its identity are one and the same; there is nothing to that self besides its defining identity. If that identity is not determinate then there is no determinate self. The ontological implications of this for the reality of selves are clearly expressed in the following quote.

[S]elves are not independently existing soul-pearls, but artefacts of the social processes that create us and, like other such artefacts, subject to sudden shifts of status.<sup>9</sup>

This position claims that selves can come and go. If this is the case, then there is nothing permanent or continuing to the self. Therefore, selves don’t exist. In the following sections, I want to address the reasons for this scepticism. I want to argue that the self is

---

<sup>7</sup> This becomes clear when we look at the example of the Ship of Theseus or that of the bronze statue of David. See R. Nozick, *Philosophical Explanations*, Oxford, Clarendon Press, 1981 for a detailed discussion of the Ship of Theseus; L Rudder-Baker discusses the relationship between the piece of bronze and the bronze statue of David ‘A metaphysics of ordinary things and why we need it,’ *Philosophy*, Vol 83, 2008, pp 5-24. She argues that constitution is not identity.

<sup>8</sup> Dennett, Chapter 13, *Consciousness Explained*, Little, Brown, Boston 1991.

<sup>9</sup> Dennett, 1991, p. 423.

not identical to its (personal) identity and that a focus on identity alone ignores an important feature of our self-phenomenology. As a consequence, that feature is left unexplained and unaccounted for. I want to demonstrate that psychological continuity, while important for individuality, is not the hallmark of a self whereas ipseity is. For there to be ipseity, there needs to be consciousness and a minimum form of subjectivity such that there is something it is like to be, either an awareness of existing or of experiencing some thing.

The general problem of personal identity can be summarised as follows: If there is such a thing as the self and it is always the same self of which one is conscious from moment to moment, then there should be something about that self that can be identified and re-identified as *that same self* despite changing circumstances and the passage of time. These are the so-called synchronic and diachronic aspects of identity; what constitutes one's identity at any given time and the subsequent identity conditions that need to be fulfilled in order for sameness/continuity of identity to be established. It has been the difficulty of providing determinate identity criteria for the self over time that has led to a general rejection of the substantive self.<sup>10</sup>

I argue that there are short-comings in the personal identity debate in that it fails to capture the self phenomenology at the heart of the Cogito and represented by Ricouer's ipseity. This phenomenology is left unexplored and unaddressed. As such, the self is not fully subsumed or accommodated under the rubric of personal identity.

## 2.4 Locke on personal identity

Locke was probably the first philosopher to focus the debate about the self on the issue of personal identity. He recognised that there was a problem about the identity of this

---

<sup>10</sup> By substantive self, I mean a self as a distinguishable substance with a physical instantiation which independently provides sameness and continuity by its perseverance through time. This is Parfit's 'further fact', 'Personal Identity' in *Reasons and Persons*, Oxford University Press, Oxford, 1986, p 225. Shoemaker argues that it is akin to a truism to claim that the self is a substance or entity. He sees 'subject' and 'substance' as synonymous. See Sydney Shoemaker, 'Are Selves Substances?' *Self-Knowledge and Self-Identity*, Cornell University Press, 1963. Armstrong claims that Hume disagreed with Locke because of his positing of substrata, a 'thin' particular in which properties inhere. Locke argued that the property of 'extension' made no sense without substance. See David Armstrong, *Universals: an opinionated introduction*, Westview Press, Colorado, 1989. I alluded to this debate in Chapter 1.

same thing, the self, particularly over time. Locke defined identity as that which is the same as itself and that which can be re-identified as the same thing both from amongst things which are not itself and in different situations over time. Hence, his definition of the 'same man' is the 'same continued life communicated to different particles of matter, as they happen successively to be united to that organised living body.'<sup>11</sup> Locke goes on to differentiate between bodies and finite intelligences. Bodies or animate substance can be identified in terms of the spatio-temporal continuity of their physical substance. Finite intelligences are to be identified with psychological continuity. Hence, a person is distinguished from a man in that a person has different identity criteria. A person, as defined by Locke, is a thinking intelligent being which can 'consider itself as itself, the same thinking thing, in different times and places; which it does by that consciousness which is inseparable from it...'<sup>12</sup> Consciousness unites personal identity by identifying itself as the same self over time. Sameness of self is the ability to be able to identify one's self with one's past actions, claiming them as one's own.

According to Locke, the idea of a person is a forensic one.<sup>13</sup> It determines legal, moral and social responsibility. In normal circumstances, a person comes into being in a socio-cultural situation. So a person is not just a token of the type homo-sapiens (the species); a person is someone who is capable of acting in a social context. Given this, it is important to know under what conditions someone *is* a person, is the same person, and can be re-identified as the same person over time. In most circumstances, this matter can be resolved simply by adverting to bodily identity or bodily continuity. However, this is not always straightforwardly obvious and is even less reliable given advents in modern medicine. Clever make-up and plastic surgery can completely change the way we look, making it impossible for even loved ones to recognise us. Less intentionally, our bodies change continuously and dramatically, from birth onwards, and we can even lose our teeth and hair and other bodily parts without this affecting our belief that we are the same person. According to Amelie Rorty we also have 'more bodily continuity than we

---

<sup>11</sup> John Locke, Bk II, Chap. Xxvii, s 8 in *An Essay Concerning Human Understanding*.

<sup>12</sup> Locke, s9.

<sup>13</sup> Locke, s 26.

can properly use'.<sup>14</sup> Our bodies are around as corpses a lot longer than we are, and a corpse is not seen as being identical to or continuous with the person it once was. So persons are not always considered identical to, nor are they necessarily identifiable with, the bodies they inhabit.<sup>15</sup> This means that a person is considered to be something other than just their body. Locke stressed that the term 'person' did not refer to nor was it identical to just the physical human body.

Human beings do appear to have an additional realm of experience that is not obviously physical, even if physically generated. Descartes claimed that the self was evidenced in our *mental* life, not in our bodily existence. One of the points he raised in his *Meditations* was that he could doubt his body and the evidence from the sensations of his body but this did not lead him to doubt himself. His 'thinking self' continued unabated. This association of the self or person with the mental realm still persists as mentioned in Chapter 1.<sup>16</sup> Locke stipulated that personal identity (human, not animal), was constituted by one's psychological or mental states. He claimed that who one is as a person is determined by the identity of one's internal mental states, not the identity of one's physical body. Within this framework, personal identity criteria for *sameness* or *continuity* of the person or self is generally considered a matter of being identified by, and consciously identifying with, one's internal mental states, including one's memories. It is the continuity of that mental life that determines the continuity of that person as the same person over time. For Locke, that continuity is provided by 'that conscious thinking thing' which experiences sensations and emotions and that is always present.<sup>17</sup> This indicates that Locke posited a *self* in addition to the person, albeit a 'thin' substance kind of self. The Lockean self appears to be the continuing consciousness or the 'I' of thought, possibly the 'subject' of experience. The person, on the other hand, is

---

14 Rorty, *The Identities of Persons*, 1976, p.9.

15 There are accounts within the personal identity literature which claim that bodily identity or continuity is what counts. However, these accounts generally use an impoverished conception of the body as a physical object whose identity criteria are the same as those for any other substance. See Jenny Teichman and Katherine Evans, 'The Existence and Identity of Persons,' *Philosophy: a beginner's guide*, 1992, Blackwell, Oxford. Also see George Rey, 'Survival' and Sydney Shoemaker, 'Embodiment and Behavior' in Amelie Rorty's *The Identities of Persons*, 1976 for a slightly more robust account.

16 See Galen Strawson's defence of a mental self in *Selves*, OUP, 2009.

17 Locke, Bk II, Chap. Xxvii, s 17 in *An Essay Concerning Human Understanding*.



something more. It is a composite of autobiographical details, interlocking memories and mental events. It is also a public persona, a social and legal entity. Self-same identity is characterised by professing sameness of self over time, remembering events as the same person and being familiar with one's own internal thoughts. Hence, Locke's imagined 'mind-swap' between the prince and the pauper leads to the prince being identified as the prince, even though he is now 'in' the pauper's body. This is because the prince is whoever he claims to be, regardless of his embodiment.

Problems with Locke's identity criteria are well known and form the basis of much of the ensuing debate on personal identity. Locke's focus was to provide identity criteria for the person that would guarantee sameness and continuity of that person. In this he is considered to have failed. His simple memory theory cannot cope with discontinuity of consciousness (e.g periods of sleep or unconsciousness), memory loss or reduplication. If the conscious self were the means of providing continuity, then it seemed to have periods where its existence lapsed. This has an impact on establishing diachronic identity of the person (and the self). The possibility of reduplication throws into doubt synchronic identity as well.

## **2.5 Hume on the non-existence of selves and persons**

Hume's now famous account of the 'no-self' was a reaction to Locke's account of personal identity. It exemplifies much of the scepticism that still permeates the modern dialogue about the existence or not of the self. According to Hume and contra Descartes, introspection does not reveal a distinct or continuous self or indeed anything vaguely self-like. All one discovers, claims Hume, is a collection of perceptions and experiences with no unity, continuity or permanence. This is illustrated by his famous passage below,

For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe anything but the perception...<sup>18</sup>

---

<sup>18</sup> David Hume, Bk I s.VI part IV, 'Of Personal Identity' in *A Treatise of Human Nature*, ed L. A. Selby-Bigge, Oxford University Press, 1985.

Hume claims that our phenomenology does not reveal a self as claimed by Descartes.<sup>19</sup> All it reveals are a disparate bunch of perceptions or thoughts or experiences that are in constant flux. This conclusion reflects Hume's general epistemology. According to Hume, our idea of substance is based on the various impressions an object creates in us. We may think something is singular, we may think it is continuous, we may think it identical to itself, but these are not things we can really know. Our judgements are based on the seeming continuity of identity over time. According to Hume, strict identity is that which 'remains invariable and uninterrupted thro' a suppos'd variation in time'.<sup>20</sup> Strict identity is therefore incompatible with change. He claims we don't *observe* continuing identity in the world around us; we develop an idea of it from relations of resemblance, causality and contiguity.<sup>21</sup> Thus, the same is true of self-perception.

He argued that the identity conditions for persons are the same as for all substances.<sup>22</sup> Hume claims that we can't know substances but substances trigger ideas in us by creating impressions. We don't perceive a self but only the collection of thoughts and perceptions which sporadically occur in our minds. These 'impressions' are discontinuous and disparate and continually in flux. There is nothing invariable over time to give us the impression of a same, continuous self. From this, Hume concludes that the self is an illusion. Personal identity, on the other hand, just consists in the collection of independent mental entities that exist moment to moment. Memory alone provides the string to tie the bundle together. He claims the self is like a commonwealth where individual members are bound by laws. Both the members and the laws can

---

<sup>19</sup> Ricoeur claimed that this failure is not due to the non-existence of the self, but to the reductive methodology and the basic assumptions inherent in such accounts. The self is not simply given to introspection as Descartes espoused. The self and its ipseity are revealed through the paradoxes of idem-identity.

<sup>20</sup> Hume, Bk I s.VI part IV.

<sup>21</sup> Hume's account is largely epistemological. It concerns what we can know about ourselves and the world, given that we only receive impressions. Robert Coburn reflects a similar despair about our inability to provide strong supporting grounds for our everyday identity claims. He concludes that either these are irreducible or we must accept that the natural view has limitations. However, this also means that ordinary objects fare no better than the self. Robert Coburn, 'Personal Identity Revisited' in *Canadian Journal of Philosophy*, vol. 15, no. 3, Sept. 1985.

<sup>22</sup> A.J. Ayer claims that Hume accepted Locke's analysis that persons were identifiable in terms of their mental states rather than their physical states, but that he used the same identity criteria for each. 'Bodies and Selves' in *Hume*, Oxford University Press, 1980, p.51.

change.<sup>23</sup> So not only are there no selves but there are no persons either. Selves do not exist at all but persons are no more than illusions created by our phenomenology.

While some philosophers have attempted to deal with Hume's scepticism, most accept the parameters of the debate as set out by both Locke and Hume. This means that discussions about the existence or not of the self tend to focus on establishing determinate identity criteria for persons by establishing what constitutes personal identity. Their lack of success has led to an increasing scepticism about the possible existence of a self, especially one that has Cartesian-like properties such as continuity or permanence. Hume's scepticism appears compelling. Here is a position that directly contradicts the phenomenology that is at the heart of the Cogito. If there isn't a Cartesian-like self then we could stop looking and focus instead on persons as the correct ontological category. Persons, at least, may be real. It is persons that act in the world, rather than bodies, and persons are not identical to human bodies. The elder Strawson argued that there are two basic particulars in the world – physical bodies and persons.<sup>24</sup> What distinguishes persons from just bodies is that persons can be individuated and identified by ascribing both physical *and* mental predicates to them.

It is one and the same person that can be ascribed the thought that it is raining and that also has the broken leg. According to the elder Strawson, 'one's states of consciousness, one's thoughts and sensations, are ascribed *to the very same thing* to which these physical characteristics, this physical situation, is ascribed.'<sup>25</sup> One cannot ignore the existence of mental predicates without ignoring something real about the world. Mental states are very much a part of what it is to be a human being; they help constitute a person. However, providing determinate identity conditions for persons is just as problematic as it is for selves. This difficulty has led to scepticism about the ontological status and existence of both.

---

<sup>23</sup> Hume, Bk I, s VI part IV.

<sup>24</sup> Peter Strawson, *Individuals*, 1956.

<sup>25</sup> Strawson, 1956, p.331.

## 2.6 Parfit and impersonal personal identity

As already mentioned, the literature on personal identity largely accepts the non-existence of selves and focuses on establishing the existence or not of persons. It is evident from that debate that establishing the existence of persons is as problematic as establishing the existence of selves. Most writers fall within two main camps, the neo-Lockeans who tend to defend the existence of persons, (not selves) illustrated by Shoemaker and Proust, both of whom I discuss below, and the Humean sceptics towards both persons and selves, such as Parfit and Dennett. I address Dennett's views in some detail in the next chapter and again in Chapter 5. Here, I will focus on Parfit. Although Parfit has been referred to as a neo-Lockean because of his adaptation of Locke's psychological continuity criterion, this 'Lockeanism' is evident only in relation to his position on personal identity criteria.<sup>26</sup> In other respects, he follows Hume's scepticism in relation to the existence of both selves and persons. Parfit's views have been particularly influential and most writers that follow him try to address the problem cases he discusses. Parfit is of particular relevance to my thesis because he is deliberately attacking any kind of entity theory in relation to the self or persons.

Parfit is addressing what he sees as three false assumptions surrounding personal identity. The first is a belief that identity rests in the separate existence of a 'core of permanence,' something continual and immutable over time, an essential being. The second is the belief that there is always a 'fact of the matter' in regards to this identity; that there is a determinate answer to the question, is this the same thing as before? The third are the moral implications of continuing identity, that there needs to be a fact of the matter in order for the subject to have moral status. He argues against all three, drawing on Hume's scepticism and revising Locke's criteria of identity. He claims there is no self; there is no person distinct from bodies and mental events; and personal identity is indeterminate and is not what matters.

---

<sup>26</sup> John Barresi and Raymond Martin refer to Parfit as a neo-Lockean, 'History as prologue: western theories of the self,' *The Oxford Handbook of the self*, ed. Shaun Gallagher, OUP, 2011, pp 41-42.

Parfit claims that there are two theories about what persons are; the Ego theory and the Bundle theory.<sup>27</sup> The Ego theory posits a continuous *subject* of experience, something that is distinct from the actual experiences. The Bundle theory claims that we cannot explain unity or continuity by adverting to a self or a person or a subject of experience. Unlike Hume, Parfit does not claim that selves and persons are illusions. However, he says they represent linguistic practices rather than what is real. Like Dennett, he claims they are logical constructions only. Similarly, the self is just a posit, used for convenience by the folk. There is no me, there are only the mental events that are ascribed to a 'me.'<sup>28</sup> One's personality or character is the external manifestation of what constitutes one's personal identity. He claims there is nothing 'over and above' identity, no extra self, no metaphysical entity or essential being, not even a Lockean-type person. Parfit argues that denying his 'reductive' thesis of what constitutes a person is to push for a 'further fact' theory which can only be the same thing as positing a separate entity, the existence of an implausible non-physical Cartesian self.<sup>29</sup>

According to Parfit, personal identity has to be defined in an impersonal way, without recourse to a 'subject' of experience, to avoid circularity. He uses the term 'event' as a primitive term that does not pre-suppose the existence of an entity in the way the term 'state' would. Consequently, one's identity through time can be described without remainder as a collection of individual mental events united by certain causal relations, Parfit's Relation R.<sup>30</sup> As claimed by Parfit,

[a] person's existence just consists in the existence of a brain and a body, and the occurrence of a series of interrelated physical and mental events.<sup>31</sup>

---

<sup>27</sup> Derek Parfit, 'Divided Minds and the Nature of Persons', *Mindwaves*, eds. Colin Blakemore & Susan Greenfield, Basil Blackwell, Oxford, 1987, pp.20-21.

<sup>28</sup> I distinguish our everyday use of me and you from Parfit's by using inverted commas. Thus you or 'you' distinguishes between the entity/non-entity assumptions.

<sup>29</sup> Parfit, 1986, p. 225.

<sup>30</sup> Parfit states that 'psychological connectedness is the holding of particular direct psychological connections. Psychological continuity is the holding of overlapping chains of strong connectedness.' Relation R is continuity and connectedness caused in the right sort of way. See, *Reasons and Persons*, p. 206.

<sup>31</sup> Parfit, 1984, p.211.

One's individual identity is composed of those events and what makes them part of one's identity is Relation R. Parfit argues that Relation R is the only thing that matters as there is no additional intrinsic self or person, other than as defined within society. Like Hume, he claims we falsely believe in a separate self or a person because we have the idea of one. The only thing that counts, he claims, is the relation between one mental event and other mental events which are caused in the right sort of way. This means one could survive one's death if there were enough causal connections between your current mental states and some other mental states at some future time. However, those future mental states could exist in another medium, along with those of others and still count as 'yours'. This means I could be said to 'survive' if enough of my mental states have the right sort of causal connections with someone else's at some future time or in some other place. Here Parfit presents various scenarios to support his case, such as brain-body swapping, tele-transporting to Mars and reduplication, all of which are designed to test our intuitions about identity. He does this to demonstrate that 1) it is Relation R that matters, not personal identity and 2) there is not always a fact of the matter about which one is 'you' and whether or not 'you' have survived. He claims there can be no determinate identity conditions for persons because there is no entity distinct from its psychological identity. Your duplicate, your tele-transported copy or your re-embodied split brain could equally count as 'you' or not 'you.' There is no fact of the matter either way.

Parfit reduces what constitutes one's personal identity to relations of psychological continuity and connectedness between mental events. He then uses this model to claim that there will not always be a determinate fact of the matter about one's continued survival. He argues against those advocates of personal identity or self who claim there must *always* be an answer to the question 'will it be me?' He claims that those who advocate this sort of determinacy are actually positing a 'non-reductive' account of personal identity. This entails, according to Parfit, that they are actually positing a further fact or Cartesian ego. He is assuming that the self as a distinct entity can only be conceptualised as some form of 'soul-pearl' or 'further fact entity.' Like Hume, he claims that there is no subject or person *in addition* to the experiences. The term 'person' just refers to the impersonal contents of the mind of a particular human body at

a particular time. This reduction, he claims, shows that what matters is not the survival of some 'ego', self or person, but the survival of enough of the right sort of connectedness between your current mental states and some future states. Continuity of identity is ephemeral but is provided to some extent, although not always convincingly, by the links between one's current mental states and the relations they stand to each other, their causal connectedness or Relation 'R'. Parfit claims that we are falsely attached to what is a nebulous self or person that has no ontological permanency, continuity or reality. Such entities do not exist. He further advises that we shouldn't attach any importance to the survival of our personal identity or psychological continuity. He claims that such a position is not just ill-placed but is destructive for the human race. Identity is not what matters.<sup>32</sup>

## 2.7 Critique of Parfit's 'no subject' personal identity

There have been numerous responses to Parfit's view on personal identity. Philosophers who do not like Parfit's conclusion that selves and persons don't exist yet who do not want to support a Cartesian, non-physical self, have either tried to establish more robust criteria for personal identity or have put forward alternative conceptions of the self.<sup>33</sup> For the latter, positing a narrative account of the self has been the most common.<sup>34</sup> I discuss the feasibility of narrative versions of the self in the following chapter, Chapter 3. Here, I want to address problems with Parfit's position. While he does raise issues about the reliability of identity criteria that need to be addressed, his account of personal identity does not demonstrate that there is no *subsequent* person or self, or the possibility that the self/person might be a complex entity rather than a simple 'soul pearl.' I addressed the ontological implications of this view briefly in Chapter 1.

Also, while we tend to accept the framing of his various scenarios and their outcomes, these are not as intuitively straightforward as they first appear. They assume that identity

---

<sup>32</sup> See Parfit, 'The unimportance of identity,' *The Oxford Handbook of the Self*, 2011.

<sup>33</sup> See for example, Joelle Proust, 'Thinking of Oneself as the Same' *Consciousness and Cognition*, Elsevier, Vol. 12, 2003, pp. 495-509. Online at [www.sciencedirect.com](http://www.sciencedirect.com); and Schechtman, *The constitution of selves*, Ithaca, Cornell University Press, 1996.

<sup>34</sup> For exemplars see Dennett, 1991; Schechtman, 1996; Blackmore, *The Meme Machine*, OUP, 1998; Velleman, *Self to Self: selected essays*, CUP, 2006.

is not bodily-related and hence is transferrable. It is my claim that lack of determinacy re identity does not necessarily mean there is no entity. Most importantly, aspects of our self-phenomenology are not accommodated within his model.

First, let's look at Parfit's Relation 'R'. Parfit claims it is one's thoughts, feelings, memories that constitute one's identity and, hence, oneself 'without remainder.' He then claims that these can be described impersonally as 'this' thought or 'this' event.<sup>35</sup> He claims that continuity of identity is provided by mental events being connected in the right kind of way (x causes y and z is remembered) and having the right kind of causal pathway (i.e. occurring in the same brain). Parfit goes on to argue that strong connectedness itself is not a criterion of identity because it does not guarantee continuity over time.<sup>36</sup> He claims personal identity is transitive whereas connectedness is not. In this way, he argues that identity is not what matters; being connected in the right way does. He then puts forward a 'wide' version of the psychological criterion that allows continuity and connection to have 'either any reliable cause or any cause.'<sup>37</sup> This allows for abnormal situations such as 'quasi' (implanted) memories and non-standard connectedness such as transplanted parts of brains. Parfit's psychological continuity allows, as he says, 'continuity to have *any* cause' (his italics).<sup>38</sup> This seems to imply that being caused or connected 'in the right way' is no longer a necessary determinant of one's personal identity.

Parfit wants there to be 'non-traditional' causation to stretch our intuitions about identity. He wants to claim that if enough of my mental events are transposed or retained in some other medium, like another brain, then 'I' could be said to survive, even if my original brain and body were destroyed. According to Parfit, because you don't exist anyway, independent of those mental events, it doesn't matter that you think it won't be you. The important part of 'you,' if there is one, is Relation R and enough of that *could*

---

<sup>35</sup> Parfit, 1984, pp.252-255.

<sup>36</sup> Parfit, 1984, pp.207-209.

<sup>37</sup> Parfit, 1984, p.208.

<sup>38</sup> Parfit, 1984, p. 208.



still hold. Parfit says that non-standard causation holds as much as standard causation.<sup>39</sup> ‘You’ would still be identity-connected to those mental events even if they were not caused in the ‘normal’ way. This means the right way has now become very wide. In fact, it appears that Parfit doesn’t have to have all my mental states or even a specific sequential set of mental states in order for it to count as ‘me,’ as his partial brain transplant scenarios indicate. This seems just wrong. How could I be said to have survived if only a randomly chosen collection of my mental states were implanted into someone or something else? Although Parfit claims that quantity does matter up to a point, as it does in the Sorites paradox, he says the question ‘will it be you?’ is nonsensical.<sup>40</sup> For Parfit there is no fact of the matter because there is no entity. Because there is no entity, no you, then identity also doesn’t matter. Parts of you surviving are as good as ‘you’ surviving – or not.

There are several issues to address here. First, it is not clear that just placing parts of my mental states into another brain, machine or in storage somewhere constitutes preserving me or even ‘me’ in any genuine sense. Parfit’s non-standard or ‘wide’ causation and connectedness leaves him open to the criticism that his connectedness criterion is loose enough to accommodate a range of unusual and seemingly insubstantial claims of continuing identity. As Coburn suggests, it could just be a matter of having a selection of my memories implanted into your brain to claim that ‘I’ am still alive, even though your memories are in there too.<sup>41</sup> Or even less convincing, it could be that I have adopted your ideas or your ideas have caused enough ideas in my brain for it to be said that ‘you’ have survived after death. Parfit, of course, doesn’t want to commit himself to how much is enough because, ultimately, he wants to claim that one’s identity doesn’t matter for survival. For him, there is no I to survive or not survive. However, the way he sets up his scenarios does beg the question. It seems that, for Parfit, Hume’s bundle is not only loosely tied together, it is not *essentially* tied together.

---

<sup>39</sup> Parfit, 1984, pp. 208-209.

<sup>40</sup> Parfit, 1984, pp. 231-233.

<sup>41</sup> Robert Coburn also argues that Parfit’s reductive analysis can mean that ‘I’ continue to survive because my mental states have had causal effects on someone else like my best friend. See ‘Personal Identity Revisited’ in *Canadian Journal of Philosophy*, Vol. 15, no. 3, 1985.

Parfit goes on to argue that memories do not assume personal identity, as some critics have argued, because they are not always caused ‘in the right way;’<sup>42</sup> they could be false or even ‘quasi’ memories, memories that are not yours but which you claim to remember because they were suggested or even implanted. I am not sure this line of argument works. It is true that memories are not always reliable, but they do seem to imply a rememberer. I think Parfit’s point is that they don’t imply *the same rememberer*, if false, thereby undermining continuity of personal identity. I’m not sure this works. If I misremember that it was me that cut my finger when I was two when in fact it was really my sister, this means that I have a false personal history. This may have had an effect on my character, but I don’t see how it disconnects memory from personal identity. I can still engage with (some of) my memories in a personal way, as the supposed original actor or experiencer of the event. For many of them, there was a time when the memory *was* the actual experience that caused the memory. There would still be continuity of identity.

Parfit also relies heavily on memory as an indicator of identity and psychological continuity. Yet not all our experiences are remembered. Some of those experiences create effects or leave traces that can influence later behaviour. It also seems like one’s memories are one’s own because they are from a particular singular and unique perspective. In response, Parfit would argue that this still reflects a Cartesian hangover, that there is no *personal* identity because there is no person. I would suggest that being a person and being a certain *kind* of person are two different things.

I want to have a short look at Parfit’s test cases. He, like Shoemaker and others, assumes that brain-swapping is identity-preserving.<sup>43</sup> He then wants to claim it will not always be determinate who has survived, showing that personal identity doesn’t matter. ‘I could be said to survive in two bodies, if my brain were cut in half, or if my information were transported to Mars and my duplicate survived on earth, allowing for the possibility of two or more ‘mes.’ Each of these scenarios would be better than death,

---

<sup>42</sup> See Albert Shalom and John Robertson, ‘Hartshorne and the problem of personal identity,’ *Process Studies*, Vol. 8, no. 3, 1978 for a case for the presupposition of a rememberer.

<sup>43</sup> Sydney Shoemaker and Richard Swinburne, *Personal Identity*, Oxford: Basil Blackwell, 1984, pp. 108-110.

even given that they can't both be 'you.' Parfit claims that commissurotomies, strokes and other accidents of nature are real-life cases of duplicate consciousnesses or half brain viability. Physiologically, these scenarios are doubtful. Left/right half brains don't work identically or like full brains and, when this happens by accident, the resulting person is much depleted. Brains also compensate and develop and adapt, depending on stimulus and circumstance. They are living organs. It is not just memory that makes one's brain one's own. It is hard to imagine how one's psychology can be readily transplanted, even if one's memories could be. Van Inwagen raises a similar concern. He claims that the view that brains are like blank computer discs awaiting informational input is nonsensical. He points out that, 'a human brain is supposed to grow to its adult size over a span of years during which it receives a certain sort of sensory input and produces all sorts of outputs that influence its environment.'<sup>44</sup> Information is not stored in a brain like it is in a computer. Brains themselves are living organisms that react to stimuli in certain ways and, as a consequence, change. He has similar concerns about the possibility of psychological information extraction and placement in another medium. He claims that materialists must show in what this information consists if not the body/brain-states of the person. While I do not agree with van Inwagen's ontological views about what can be said to exist, he does have a point in relation to the plausibility of brain transfers, cloning and teletransportation.

One could agree with Parfit that one's identity is constituted by one's thoughts, feelings and memories. We might even say that that is what constitutes a person. However, Parfit wants to prevent us from positing a person because he claims this would slip us into positing an *additional* entity or object that has those properties. I am not sure why a person so constituted would not count as an entity, just as any complex entity could be said to be constituted by its properties.

---

<sup>44</sup> Peter van Inwagen, 'Materialism and the psychological-continuity account of personal identity,' *Philosophical Perspectives, 11, Mind, Causation, and World*, 1997.

As mentioned in Chapter 1, Parfit reflects a certain metaphysical approach to entities, reminiscent of trope theorists.<sup>45</sup> Objects are nothing more than collections of properties or tropes. You get objects out of ‘compresent’ properties and property-relations. As I mention below with Hume, this could lead to the possibility of uninstantiated universals or particulars.<sup>46</sup> Parfit’s reduction of persons seems to take a similar turn. The properties, i.e. mental events, exist but not the object that possesses them. If an object is just a bundle of tropes, then tropes (property particulars) are ‘object-like entities’ that have their own identity-conditions. This also leads to an infinite regress.<sup>47</sup> Like Heil, I would suggest that mental events and other such properties may be better conceived as ‘modes of being’ which cannot be separated from their object, in this case the person (or self). If properties have grounds for entity existence then so would the objects in which they inhere. If not, selves and persons are no more nor less existent than any other object constituted by its properties. This situation would just reflect the ontology of the world in which we live.

It is by no means clear that removing the personal from the description of mental events does the work Parfit claims. Parfit claims that we only need to reintroduce the term ‘person’ when discussing the *content* of the thoughts. Mental events have content and one presumes it is through that content that they play a role in the cognitive system, and to be classified variously as intentions, beliefs, memories, feelings. If content cannot be understood *impersonally*, then I am not convinced that mental events can be either. The same applies to the Lichtenberg critique of Descartes.<sup>48</sup> Even if one can say of another person that there is thought rather than a thinker, at some point there has to be knowledge or awareness that a thought has occurred, in the observer if not in the mind

---

<sup>45</sup> John Heil engages in a discussion about trope theorists, properties and objects in Chapters 8-10, *from an Ontological Point of View*, Clarendon Press, Oxford, 2003. It seems properties can be viewed as either particulars or universals. If viewed as universals, the problem is the nature of their existence. Are there real existing abstract entities like properties and other general terms? Heil argues for modes rather than tropes. Armstrong argues for the existence of universals, not as existing outside space and time in a separate realm but as real abstract entities. He claims one cannot make sense of laws of nature without recourse to universals; ‘How do particulars stand to universals?’ in *Oxford Studies in Metaphysics: Vol. 1*, (ed) Dean Zimmerman, OUP, 2004.

<sup>46</sup> E.J. Lowe, *The Four-Category Ontology*, Clarendon Press, Oxford, 2006, p.72.

<sup>47</sup> Heil, 2003 p.141.

<sup>48</sup> Parfit refers to Lichtenberg in detail in *Reasons and Persons*, 1984, pp. 225-227.

that has the thought. Who or what has that knowledge or awareness if there is no person or self? This is reflective of the limitations of a functionalist description of cognition from which consciousness is excluded, and seems superfluous to, explanation. I discuss this problem below.

Parfit claims that if you don't accept his reductionist thesis, you are positing an additional (Cartesian) entity. I think this is a false dilemma. He claims either you accept his thesis or posit an immaterial entity. This does not need to be the case, given there could be alternative conceptions of an object and its relation to its properties. One needn't be a nominalist. At the same time, Parfit is working from only one concept of a self or person, the inner immutable substance. Reduction for Parfit is not a (better) explanation of macro-properties or things. It is not an intertheoretic reduction.<sup>49</sup> He does not say that a person or self is reducible to or really is Relation R. Reduction for Parfit makes the original phenomenon disappear. There is Relation R but no self or person. I return to this point in Chapter 5 in relation to physicalist explanations of the self and reduction.

Parfit claims that there is not always a fact of the matter about one's continuing identity, hence there is no entity. This may be true in some cases. It is likely to be true in several of Parfit's brain-swap examples. I am not sure that this entails what Parfit claims, that the first person experience is just an illusion or a logical construct. To highlight some of the problems, I refer to Sydney Shoemaker's argument against Parfit.<sup>50</sup>

One common idea in relation to identity of the self/person is that the self or person is analogous to a commonwealth or a club. This means that individual mental events are akin to the members of that club. Shoemaker claims that this type of reduction assumes that, just as the members of a club can exist independently of that club, then mental

---

<sup>49</sup> I discuss Patricia Churchland's account of the relationship between theory and reduction in Chapter 5 when discussing Dennett and Metzinger. *Neurophilosophy: towards a unified science of mind-brain*, MIT Press, Mass., 1986.

<sup>50</sup> Sydney Shoemaker, 'Parfit on Identity', *Reading Parfit*, ed. Jonathan Dancy, Blackwell Publishers, Oxford, 1997, pp. 135-148.

entities can exist independently of there being persons.<sup>51</sup> Hume actually says that perceptions ‘...may exist separately, and have no need of anything to support their existence.’<sup>52</sup> If true, it is difficult to see what kind of identity criteria such mental entities could have in order for them to qualify as a particular mental event. A perception of a blue room, for example, would need some identifying criteria to make it simultaneously a perception of blue and of a room.<sup>53</sup> Those criteria usually entail that they be the experiences or perceptions of something, in order for such mental entities to qualify as being perceptions or experiences or even mental. Shoemaker claims that it is unlikely that any such independently existing Humean mental entities actually exist. Parfit’s insistence on removing the personal from his definition of identity leaves him open to the same charge.<sup>54</sup> In Parfit’s account, it is almost coincidental that particular mental events occur in this brain rather than another brain or even in another medium. There is nothing in his account that makes them Sandra’s or which even makes them mental as it is commonly understood. Against Parfit, Shoemaker argues that experiences are adjectival on *experiencers* in the same way dents are adjectival on a surface. This is how they are defined. That is what defines a person. Shoemaker also claims that this is not just a matter of language use. Defining the mental *as* the mental means identifying it as ‘a particular mental kind.’ Something is recognised as a belief or a perception by its being related to a larger system in which it plays a causal role. Its causal role will be determined by the meaning or significance it has for the personal system in which it inheres. Shoemaker claims that this means there is ‘a necessary ontological dependence of experiences on the existence of persons or other mental subjects,’ such as selves.<sup>55</sup>

---

<sup>51</sup> Shoemaker, ‘Parfit on Identity’, 1997, pp. 138-140.

<sup>52</sup> Hume, Bk I s VI Part IV.

<sup>53</sup> Some would argue that if the sensation were more basic such as a sensation or perception just of ‘blue’ then it is not obvious it needs to belong to someone rather than just be raw input data. I am not convinced that we ever experience raw data in an undifferentiated way even at a pre-conscious level. There is evidence that sense data is differentiated at the site of contact before it is sent to the brain for further processing. See *Reasons for Realism: selected essays of J.J. Gibson*, Edward Reed, and Rebecca Jones, (eds), Lawrence Erlbaum Assoc., Hillsdale, N.J, 1982 and Timo Jarvilehto, ‘Efferent Influences on Receptors in Knowledge Formation’, *Psychology*, 1998, 9, Issue 41.

<sup>54</sup> Peter Strawson does the same thing in *Individuals*.

<sup>55</sup> Shoemaker, ‘Parfit on Identity’ 1997, p. 139. For a similar criticism against Hume see Harold Noonan, Chapter 4, ‘Hume’ in *Personal Identity*, Routledge, London, 1989.

Functionalist descriptions of perception are also impersonal and do not rely on positing persons or selves to explain perception. However, such accounts have been criticised for leaving out the experiential side of perception. This is often referred to as the explanatory gap. Functionalist accounts contribute little to an explanation of the self phenomenology and often have to resort to positing such phenomenology as an illusion as I discuss in Chapter 5. A standard critique of functionalism has been its neglect of qualia and the role of consciousness in action.<sup>56</sup> Parfit just ignores the phenomenology. In all his thought experiments, he never once thinks what it must be like from the perspective of the person undergoing the experiment. He wants to show that personal identity is not important and should not be what matters in cases of one's supposed death or survival.

The Sorites example is a case in point.<sup>57</sup> In this, he assumes that which he is attempting to describe, the impersonal. He claims that our intuitions will not be able to tell when or at what point of gradual cell replacement we will cease to exist or will become someone else. This is probably true. However, were it to happen in real life, it is very likely there would be a fact of the matter. There would likely be a point at which the person undergoing the surgery would no longer be anyone in particular. While this demonstrates that it is possible to change one's identity or to even destroy the person one once was, this doesn't demonstrate there never was a person in the first place, that there never was a determinate identity. Again, similar arguments can be used against the existence of all objects. Unger, for example, used a Sorites argument he credits to Eubulides to repeatedly put forward the position that there are no objects in the world, that nothing exists, 'not rocks, not desks, not plants, not humans.'<sup>58</sup> He has since recanted this extremist position but it illustrates my point that even if the person or self

---

<sup>56</sup> There are many writers on this topic but see John Searle, 'Minds, Brains and Programs,' *Behavioral and Brain Sciences*, vol.3, 1980; Thomas Nagel, '[What is it like to be a bat?](#)', *The Philosophical Review* LXXXIII, 4, 1974: 435-50, and Ned Block, 'Troubles With Functionalism,' *Readings In Philosophy Of Psychology*, ed Ned Block, Cambridge, Harvard University Press, 1980a. [http://www.mcps.umn.edu/philosophy/9\\_12Block.pdf](http://www.mcps.umn.edu/philosophy/9_12Block.pdf).

<sup>57</sup> Parfit, 1984, pp. 232-234.

<sup>58</sup> Peter Unger, 'There Are No Ordinary Things', *Synthese*, Vol. 41, 1979, pp 117-54; 'I Do Not Exist' in *Perception and Identity*, ed. G. F. MacDonald, London: Macmillan, 1979, pp. 235-51; and 'Why There Are No People,' *Midwest Studies in Philosophy*, Vol. 4, 1979, pp 177-222.

is vulnerable, its identity is not always determinate. These are not just issues for the self or persons; to some extent, these are issues for all objects.<sup>59</sup>

While we may be able to imagine situations where our intuitions about continuing identity are stretched to the point of non-committal, this does not mean that we can readily accept that thoughts, feelings, emotions or perceptions are only incidentally tied to their host. Not only do these internal events take their identity as internal mental events from their content, as Shoemaker points out, the mental events that occur in any one mind/brain are not completely distinct separable items in the way that Hume and Parfit claim. They are part of a particular kind of continuous, historical, spatial network. The type of emotional response one has to a situation or the quality of a perception will be largely determined by precedence, individual circumstance and current situation. These will be unique and very personal and will determine what mental events could occur in any one system and at any given time.

Parfit himself acknowledges that we care about our lives and our lives have meaning to us.<sup>60</sup> It is not coincidental that my thoughts are mine with their particular content and that yours are yours with their particular content. The precise content will not just be determined by one's perceptual experiences but by one's current state of mind and the kind of events that have happened in the past. This is what is meant by 'personal.' We can remove talk of the personal but it does seem to remove an aspect of human life that is fundamental to the way we relate to each other and ourselves. This is Ricoeur's point about the ahistoricity of the traditional Cartesian conception of the self. He argues that, while we can generalise about certain traits or capacities, ultimately a person is a *particular* individual person, and a self is always a particular individual self. Ricoeur claims you can't remove the historical context or the specific embodiment from the person. McDowell raises a similar criticism against Descartes which is equally relevant to Parfit. If consciousness becomes devoid of its human and particular context, then it becomes detached and re-ified as a featureless entity. McDowell argues that this has

---

<sup>59</sup> See P. Unger, 'The problem of the many' in Zimmerman (ed) *Oxford Studies in Metaphysics: Volume I*, Clarendon Press, Oxford, 2004 for a less extreme position.

<sup>60</sup> Parfit, of course, wants to make us care less. He is arguing that personal identity (which includes persons and selves) doesn't matter.



made reductionist accounts of identity possible because once the Cartesian entity is removed, all that is left are the 'diverse actions' mentioned above. It then seems that personal identity can be fully accounted for by a reduction to this mental collection. He suggests that continuity of consciousness (Locke's self) can be viewed as just an alternative way we keep track of the continuing human being, subjectively or 'from the inside'.<sup>61</sup>

Of course, Parfit could claim that Shoemaker's network of mental events is the same as his Relation R as it amounts to no more than having the right kind of causal history. In most 'normal' circumstances, your thoughts are yours because they have the right historical connectivity. In fact, if enough of your thoughts were transferred to or existed in someone else, Parfit would claim that 'you' have survived in them. My mother held a somewhat similar view of immortality. She talked of surviving 'in' her children, not just through the continuation of her genes but by the continuation of her ideas, values and beliefs in her children's minds. According to Parfit, this is a very real form of immortality. If enough of her beliefs have survived, there is the right kind of connectivity and there is no other closer continuer that could qualify as my mother, then I (or any one of my sisters) could be my mother. However, this highlights a major objection to Parfit's account. My mother did not believe that she, as a *conscious subjectivity*, would really continue on past her death. Nor did she believe that she could survive in my or my sisters' brains, no matter how many of her mental states we shared. It is the memory of her and her ideas, not *her memories* that live on, and that is an important distinction. More importantly, none of us share our conscious existence with our mother's consciousness. Even while I may share many of her beliefs, and have a lot of knowledge of her life, I cannot engage with her memories, even as quasi memories. My experience of her mental life is qualitatively different from my experience of my own mental life.

---

<sup>61</sup> McDowell is defending a neo-Lockean approach to personal identity. McDowell, 'Reductionism and the First Person' in *Reading Parfit*, ed. Jonathan Dancy, Blackwell Publishers, Oxford, 1997, p.233.

## 2.8 Personal identity and Ipseity

In the beginning of this chapter, I mentioned Ricouer and the property of ipseity. As discussed, ipseity is described as the sense of being or existence that Ricouer claimed was the essence of the Cogito. This is not to be confused with the ‘inner touch’ or coenaesthesia, which I discuss in a later chapter of the thesis.<sup>62</sup> Ipseity here refers to the experience of oneself as a thinking thing, the subjective experience of being a particular cognising entity, what I define as the sense of one’s self or one’s subjectivity. As far as we know, all conscious humans experience themselves as a conscious entity and believe themselves to be capable of self-reflection and self-exploration. We appear to introspect, and we do this from a position of selfness.<sup>63</sup> Whether it has determinate identity criteria or not, this self-entity is what Hume directed towards his own mind but which he then failed to discover. It is this subjectivity, the experience of one’s thoughts and experiences *from the inside* which is completely missing from Parfit’s account. He claims that positing such a subject entails something additional to its contents. This would give us a false belief in a separate self or person.

As already mentioned, this position is reminiscent of nominalists like Locke and the trope theorists. If one holds that objects are just collections of properties, it is not surprising that the self or person will likewise be reduced to just its properties. On the other hand, it is by no means obvious that a belief in a self or person is erroneous. As Shoemaker and others point out, an experience or perception entails subjectivity. One can engage with or ‘be’ in one’s *own* mental and bodily states in ways one can’t with just any mental or physical events. There does appear to be a genuine difference in the way we perceive our *own* internal events and the way we perceive other events in the world. We seem to be directly aware of our own internal states in a way that we are not aware of those of other people or other events external to us. Of course, our

---

<sup>62</sup> Heller-Roazen claims the inner touch is not the same as inner sense or mental awareness or consciousness. I discuss his position in some detail in Chapter 4. Daniel Heller-Roazen, *The Inner Touch: archaeology of a sensation*, Zone Books, New York, 2007.

<sup>63</sup> I am not going to discuss whether or not we really have this capacity but it is often taken as the hallmark of self-reflection. But see Murphy for an argument against introspection as a genuine capacity. Graham Murphy, ‘Why there is no such thing as introspection’, paper presented to research seminar Adelaide University, June 2009.

phenomenology could be flawed or just plain false. Dennett would argue that this ‘experiential’ difference is just the difference between receiving information one way rather than another.<sup>64</sup> This may be true. Nevertheless, it manifests as a real qualitative difference and it is that difference which separates us from thermostats and other automata. As far as we know, they do not have experiences and they do not self-reflect. This means that the phenomenon cannot be simply ignored as though it is non-existent. At a minimum, the explanatory gap should be addressed. Parfit does not do this. He does not offer an explanation for the discrepancy between what he says is the case and what we seem to experience to be the case. This makes his account not only counter-intuitive, but less plausible as well.

There is something more to ‘x’ being my thought or my perception than just the brute fact of it occurring in my brain. It does seem to belong to me or be ‘mine’ in a way that cannot be easily dismissed as mere illusion. This is in addition to the experience of subjectivity. For example, it is part of the character of perception that, in order for it to be recognised as a perception, it is located as coming from a particular ‘point of view.’ This means the perception includes information about the body, the human being, or even the person that ‘perceives’ it, depending on the level of explanation. What appears to link seemingly single individual (subjective) perceptions and what may provide the sense of continuity between those perceptions is that they are perceived from a particular point of view, within a particular body and from the *same* point of view within the *same* body (or mind) over time.<sup>65</sup> They are about ‘me’ and by ‘me,’ whether this is called the person or the self.

To push this point more strongly, there is evidence that perception is not a simple passive process of undifferentiated raw input and neuronal response. According to Gibson, it is an *interaction* between the environment and the organism.<sup>66</sup> The body’s

---

<sup>64</sup> Dennett, 1991.

<sup>65</sup> This occurrence of states within the one system is sometimes referred to as boundedness. See Glenn Carruthers, ‘A model of the synchronic self’ in *Consciousness and Cognition*, 16, pp. 533-550, 2007 for an example of this use. I use this term in a similar way and with a similar meaning throughout this thesis.

<sup>66</sup> Gibson claims that the environment provides affordances to the body, in line with the type of body, its height, position, etc. Light rays reflect off surfaces into the eye that indicate dimension. *Reasons for*

reaction to stimuli informs the experiencing subject about its own capacities and its current state. This means that it is always self-identifying and self-informing in some way. In this sense, it cannot be seen as something independent of the perceiver. It isn't just an unattached or un-owned thought or perception (as in not belonging to someone) that could, in principle, belong to someone else. The individual content of the mental state, the quality of the perceptual experience and their unique perspective will be determined by the history, circumstance and embodiment of the individual human in which they inhere. This is more than just the right causal path. It relates to the content as well. In fact, the exact same mental events could not have occurred in any other human system. They will be unique to that system.

To conclude, Parfit's (and Hume's) impersonal account of personal identity does not work because it ignores a real aspect of our phenomenology which is the cornerstone of our belief in sentience – subjectivity or ipseity and its corollary, self-consciousness. Parfit does not successfully show that the person or the self does not exist. Subjectivity does not necessarily imply a self but it does require an explanation of the experience of or sense of self. Parfit cannot just say this is a mistake and we would be better off without the idea of a self. The idea of a self is posited because of that sense of self.<sup>67</sup>

Personal identity on its own may provide criteria for individuality such as the personality or character of an individual but it is not sufficient to provide the sense of self or selfness. I may hate football, be good at crosswords and visit my friends on weekends, but this just describes my nature and my personal history; it defines what kind of person I am or have become. It does not capture my sense of self or why I feel like a 'me' from the inside. Personal identity is not identical to selfhood. What we need to posit is some kind of structure or entity that explains the experiential quality and the sense of agency. We need to posit a subject.

---

*Realism: selected essays of J.J.Gibson*, 1982. See also Timo Jarvilehto, 'Efferent Influences on Receptors in Knowledge Formation', *Psychology*, 1998.

<sup>67</sup> I argue for this in some detail in *Dennett and the self*, MA, Flinders University, 2002.

## 2.9 Psychological continuity and the problem of identity

Parfit argues that there is no person in addition to one's personal identity and says that identity is constituted from impersonal events. He explores the problem of providing determinate identity criteria for persons with his scenarios on brain-swapping and teletransportation. The resulting indeterminate conclusions lead him to conclude that there is not always a fact of the matter about who you are and, therefore, we should let go of our focus on self-interest.<sup>68</sup> I address these issues in the final section of this chapter. I have argued that you can't remove the personal from the debate without denying an integral part of the phenomenology. To do this is to ignore an important characteristic or property that serves to create the sense of self. This subjectivity or self-consciousness is what, for Locke, provided the continuity of identity over time.<sup>69</sup> However, positing a conscious self does not solve the problem of diachronic identity. When identifying a distinct self or person, there is the issue of establishing a determinate identity for that self or person; knowing that this is one and the same self or person rather than another and that one can be re-identified as the same self or person in different circumstances and at different times. Persons change over time, whether considered as purely physical beings or as constituted by a combination of physical and psychological properties.

If selves or persons do exist, then there should be some identity criteria one can use to identify a particular person as that person and to re-identify them again over time and change. There is an extensive debate on object identity and the difficulties of re-identifying objects as the same object. Some of these issues were raised in Chapter 1. The problematic case of the 'Ship of Theseus' captures many of the issues surrounding identity. It represents a supposed analogous scenario with human (physical) identity in that the original ship changes over time by having its parts slowly replaced by new ones. Disagreements arise when one has to decide whether or not the Ship of Theseus is still

---

<sup>68</sup> Derek Parfit, 'Personal Identity and Morality' *Reasons and Persons*, 1984, chapter 15. Regard for self is not due purely to a belief in a separate 'ego' self. Dennett's biological self is 'self-interested.' One could also view oneself as excessively important in the social world based on one's status, regardless of a belief in a metaphysical self or person.

<sup>69</sup> Locke, Bk II Chap. XXVII s 17.

the same (original) ship and at which point it is no longer clear that it is still the same ship. This becomes more challenging if the original parts that were replaced are reconstructed into another ship so that there appear to be two ships qualifying for the title of the Ship of Theseus. According to the definition of identity, there cannot be two contenders for the same identity. Identity is numerical and is 1:1, being identical with oneself. Each object is unique in occupying a specific spatio-temporal location or having a clearly identifiable spatio-temporal property. However, identity over time requires some kind of tracking of that physical entity through space along with specifying certain kinds of properties that can be identified as belonging to that particular entity which make it unique to itself. If there is more than one contender for the same identity then there should be some determinate way to distinguish the real entity from its close copy. If there is no determinate way to distinguish one entity from another or no clear answer to the question 'is it still A?' then there is no determinate identity. Either the determinate identity criteria don't apply or there can be no determinate identity criteria that work in all cases.

Nozick proffers a solution to this problem by introducing the category of closest continuer.<sup>70</sup> He uses the Ship of Theseus to argue that our intuitions about identity are correctly represented by the closest continuer criterion. He says that if we were to watch the progressive rebuilding of the ship of Theseus, we would claim that the ship is the same throughout that period of progressive change. However, this conclusion might be challenged if we were confronted by another ship constructed from the original materials. The ships are not identical to each other but each has a strong claim to being the original ship. Once we have two objects vying for the same identity, the issue is no longer simple. Yet prior to the appearance of the second ship, we would have had no hesitation in claiming that the ship of Theseus was the reconstructed one as its closest continuer. So it is only when we are challenged by a competitor that the determinants of identity are questioned. Nozick claims that, rather than there being some independent criterion that will determine which ship is the real ship of Theseus, in practice different people may select different candidates as the closest continuer, based on their weightings

---

<sup>70</sup> Robert Nozick, *Philosophical Explanations*, Harvard University Press, Cambridge, 1981.

of relevant factors. This means for Nozick, as for Parfit, in some circumstances there is no determinate fact of the matter as to which ship is the true Ship of Theseus or which is the real closest continuer.

In this example, Nozick shows that there can be several different factors which constitute a thing's identity but not which of these factors are necessary for determining an outcome. For instance, in the above example one contender has identical parts to the original ship but is spatio-temporally discontinuous, while the other has spatio-temporal continuity and consistent causal relations with the original but no sharing of parts. According to Nozick, a person's decision about which ship is the original will depend on the different weightings one places on the determinants of sameness of identity, ie whether one weights spatio-temporal continuity or sameness of material or some other factor like a causal pathway.

Nozick concludes that spatio-temporal continuity may be a factor for re-identification but it is not a necessary condition. Sameness of parts may also be a factor but not a necessary condition. This is particularly important when dealing with the identity and re-identification of persons or selves. Given that our cells (true of all organic or living matter) continue to multiply, die and regenerate as we age, then we cannot claim that we have the same parts as we did at a previous time. So a future self or person will be constituted by different matter. The same is true to some extent of all matter but it is especially the case with living organisms.

The metamorphosis of an acorn into a small sapling that turns into a tree involves an even more drastic change, as does a tadpole into a frog or a caterpillar into a butterfly. This kind of growth and change pose specific problems for finding criteria of identity and impact on claims about continuing identity of changing things. Humans fall into this category. One could posit that physical spatio-temporal continuity is a necessary condition for diachronic identity. This would yield a determinate fact of the matter about the Ship of Theseus discussed above. The continuing ship would be the original ship while the ship using the original material would not.

However, Nozick claims that should something happen to the continuing ship then the replacement ship would be considered the ship of Theseus instead. He says this is okay *because* we use a closest continuer criterion of identity when making judgements about things. Such a judgement is not fixed by one criterion alone that is considered necessary. And there appears to be no criteria that are sufficient in themselves. But it would seem that what is necessary, if not sufficient, for Nozick's closest continuer is some kind of historic or causal relations, similar to Parfit's Relation R. By using his account, 'you' as the same person or self are not re-identified at a different time but what he calls your 'closest continuer' is. In most cases, this will be like re-identifying oneself but, in effect, whatever constitutes your closest continuer counts as 'you.'

The reason Nozick does not think spatio-temporal continuity is a necessary condition is because one could imagine a world where things or beings flickered in and out of existence on a regular basis. He claims that in such a world we may re-identify something as the same despite not being able to track it continually through time. While this sounds feasible, I am not sure this is the case. In such a world, we may develop a totally different concept of identity whereby we accept that there is no guarantee that this object appearing now is the same as the object that appeared ten minutes or ten seconds ago. In fact, in such a world we may not be that bothered about issues of identity through time because it would be irrelevant. It is not definite that a chair, for instance, that came and went continually would be classified as the same chair unless one could develop *an alternative criterion that was necessary for identification*. This could be sameness of parts, but there is no independent criteria one could use to demonstrate without doubt that the parts were the same, even if each item were uniquely branded. In such a flickering world, things have the same kind of existence as waves on an ocean. They come and go regularly but there is no way of claiming that this or that is the same wave. Re-identification of physical objects would be so difficult that it is likely that strict identity conditions may not exist in such a world. More importantly though, the lack of determinate identity in such a world is the same for all objects. The self is no more nor less re-identifiable than any other object. Just as the Ship of Theseus illustrates the difficulty of re-identifying any object as the same through time and change, the



flickering world does likewise. It is important to keep in mind that diachronic identity is a problem for all objects, not just the self or persons.

The possibility of a flickering world forces our claims of identification and re-identification of persons to be other than spatio-temporal continuity or sameness of parts. If such a world had sentient beings capable of thinking and talking about themselves, then such a being could *claim* to be the same person who appeared five minutes ago and might, as proof, pick up a conversation where they left off or otherwise indicate continuity through some other means, like one's behavioural characteristics or one's recollections. Their continuing identity would be dependant on some kind of psychological continuity. However, rather like our moments of unconsciousness, it is not strictly continuous as there would be known, observable gaps during the times when this person flickered out of existence. This merely highlights the fact that although there is an alternative set of criteria for persons, psychological continuity, it is still problematic. Again, continuity through time is a problem for all objects.

Positing the 'closest continuer' does not allay the existing problem of providing determinate identity criteria. It adds little to Parfit's discussions. The only difference is that Nozick claims this is the criterion we use. We judge cases where personal, rather than physical, identity is challenged, based on which body/person/thing we think most closely ties with or continues that body/person/thing. If there is a dilemma or it cannot be easily decided, a person's decision will be based on the one that is judged to have the most or closest connections to the original. So if you were to be duplicated, your clone would not be considered to be you unless you subsequently died and it then qualified as your closest continuer.

I think Nozick is just wrong here. I doubt that I will decide my own continuity based on the closeness of some clone contender, even if I use that criterion for inanimate objects or even other people (although even in that case this is by no means certain). The interesting thing about people is we can make these decisions about ourselves, our identity, for ourselves, regardless of the objective criteria. I can claim to be me now and also claim to be the same person who visited last week and who went to that school

when young. If you present a scenario to me in which there are several contenders for being my closest continuer, I may decide on one of them despite what has been generally agreed upon as important criteria for continuity. I may even believe that none of them will be me. In fact, as Wilkes has argued, I don't care whether or not you say to me that, even though I can't survive the operation, there is an identical clone ready to be animated the moment I die.<sup>71</sup> It is no consolation to me if another person either is identical to me in terms of Relation R, or claims to be me, or will be considered me at some future time, if 'I' am faced with my own impending death. What counts for me is whether or not 'I' as this currently-experienced subjectivity will still be around. It doesn't matter that you tell me that I won't die because this clone is my closest continuer and, hence, will be me. I do not expect to wake up inside the clone's body in the same way I wake up from an anaesthetic. Of course, if that turned out to be the case then I would have to claim that 'I' had really survived. But it is highly unlikely that the subjective 'me' distributed inside my brain somewhere can be transported from one body to another and survive. Again, this possibility relies on there being nothing more to me than a series of mental events and ignores me as a perceiving subject. Van Inwagen thinks all Parfit's hypotheticals are suspect, given that information in the brain is not stored in the same way it is in a computer.<sup>72</sup> At the same time, he doesn't believe in selves as independent entities either, but he does claim these kinds of thought experiments ignore the science and are anti-materialist. Whatever the potential reality of such experiments, it is my contention that being the *same person* is different to being the *same subjectivity*.

Nozick points out that, in puzzle cases, different people may give a different response to the question 'which clone would be you?' In some cases there may be no fact of the

---

<sup>71</sup> Kathleen Wilkes, 'Multiple Personality and Personal Identity' in *British Journal for the Philosophy of Science*, Vol. 32, 1981, p.343.

<sup>72</sup> P. Van Inwagen is skeptical of brain-swapping, cloning or other duplication methods as being 'identity preserving,' given the complexities of brains and their disanalogies with computer discs. Information is not just stored in brains in complex ways but changes the structure of the brain in the process and over time. 'Materialism and the psychological-continuity account of personal identity,' *Philosophical Perspectives*, 11, *Mind, Causation, and World*, 1997.

matter, particularly when there is more than one contender.<sup>73</sup> According to Parfit, this indeterminacy means that there is nothing determinate about one's identity and, therefore, there is no unique person. He uses this to argue against the idea of a self. I think there are several issues at stake here. First, as we can see with the Ship of Theseus, there are problems with providing determinate identity conditions for all objects at some level of description. Pushed far enough, our intuitions about identity do not always yield a definite answer. There will be some cases where we say we don't know which, if anything, counts as the original object. As set out above, spatio-temporal continuity is generally considered to be both a necessary and mostly sufficient condition in most cases. This view falls out of our knowledge of living in a particular world where it appears that objects have some kind of stability, abide by the laws of physics and continue to exist independent of our immediate perception of them. Objects remain where we left them unless there has been some intervention from something else. Their physical construction and their location are usually constant enough for us to re-identify them as the same thing. While there are problem cases, as with the Ship of Theseus, we don't use this fact to doubt the existence of all things. The fact that a theory of identity can always be stretched to include cases where we may not be able to give a 100% definitive answer does not lead (nor should it lead) to real world scepticism. Likewise, the fact that objects can be changed so much they no longer bear any resemblance to the original, or can be destroyed, does not cast doubt on their once existing as a particular individual entity. There once *was* a fact of the matter. The same is true of persons. It would be demanding too much from a theory of personal identity to expect it to do more. The criteria for identity of persons or selves need not be any more robust than that for a chair or a tree. All are problematic in some circumstances.

As stated above, a unique feature of humans is that they can identify themselves without recourse to criterion of identity. They just experience themselves as a certain subject

---

<sup>73</sup> Bernard Williams also shows that our intuitions may vary depending how the scenario is presented. He claims that in a brain/body swap we cannot know if we will wake up and claim to be Brown or whether we will merely claim to remember being Brown. The usefulness of thought experiments like these has come under fire with claims that the personal identity debate is not progressing. See Williams, 'The Self and the Future' in *Personal Identity*, ed. John Perry, 1975.

with certain properties. They identify themselves accurately even when those properties change. Thus, when confronted by problem cases, they may yield different responses to each other depending on what they feel will be their future self. Pace Parfit, this does not mean there is no such thing as a person or self. In some cases it may be that there really is no fact of the matter which person you will be; it may be indeterminate like the Ship of Theseus. In others, the person or self you were has really ceased to exist. This does not inevitably lead to the conclusion that there never was a person in the first place. Just as I can blow up my house, I can destroy myself. It doesn't mean there wasn't a fact of the matter about either entities' identity prior to that incident. It doesn't mean that there never were persons or selves at some point in time.

## **2.10 Identity criteria for persons**

There have been attempts to establish identity criteria for persons that do not suffer from Parfit's indeterminacy or which don't lead to sceptical claims about the existence of persons. In 'Thinking of Oneself as the Same,' Proust revises Locke and others' 'simple memory criteria' of personal identity to provide a more robust account that is not susceptible to the criticism of circularity and which takes into account the additional capacity of co-referring.<sup>74</sup> At the same time, she wants to avoid the possibility of reduplication. She wants to give an account of individuality or uniqueness so that there *is* a fact of the matter about who I am as distinct from other possible 'me-s'. To deal with Parfit's bundle theory, she puts forward the claim that 'a person cannot exist aside from a historical process' and that there needs to be a cognitive sequence for personal identity to emerge. She claims that a self-moment is not enough for existence. A self needs recognition to know it is the same self, which implies some kind of existence through time. It should be noted here that by 'self', Proust is not referring to the Cartesian self or my self but rather to the act of self-referral. Self, personal identity and person are used interchangeably throughout her paper. Her focus is on persons, not selves.

---

<sup>74</sup> Proust, 2003, pp. 495-509.

Proust claims that the self becomes a ‘dynamic function’ of the various needs of the system to keep track, monitor and adjust to a changing environment. This, she believes, is a universal imperative, not susceptible to cultural variation. The self is constituted through this temporal process. Thus, continuity is inbuilt in the system by the actions of those self-identifying processes or mechanisms.<sup>75</sup> She goes on to say that ‘only mental agents may qualify for selfhood.’<sup>76</sup> If someone cannot reflect and adapt their actions in the light of experience or knowledge, then they do not develop into persons.

They just act according to the dictates of their current mental states or their dispositions. Animals and wantons fall into this category. According to Proust, ‘self’ refers to ‘an endogenous individual structure of the will based on a form of metacognitive memory.’<sup>77</sup> She then sets out three criteria by which a mental agent becomes, or qualifies as being, a self –metacognition, memories and the capacity for revision based on these. Ultimately she claims ‘a person is a system of dispositions, socially encouraged and trained, designed to revise beliefs, desires, intentions, and thereby become the actor/goal/target of one’s own life.’<sup>78</sup>

Proust’s account is interesting in that she tries to go beyond memory to provide continuity by stipulating the needs of the system to keep track of itself. There is some sense in this as it would seem an essential feature for one’s survival. For Proust, the ‘self’-function is represented by the capacities for metacognition and revision, which are universal characteristics of humans.

This seems to imply they are cognitively or biologically driven, rather than psychological. If so, this would make identity criteria tied to one’s physical instantiation as a particular kind of entity. While this is not a problem in itself, and is a position I will argue for myself, it moves away from the psychological criteria she seemed to be defending. At the same time, it seems to leave the ‘person’ as a set of socially acquired

---

<sup>75</sup> This is a little like Dennett’s self-representational ‘blip’ and Metzinger’s phenomenal self model which I discuss in Chapter 5.

<sup>76</sup> Proust, 2003, p.501.

<sup>77</sup> Proust, 2003, p.501.

<sup>78</sup> Proust, 2003, p.504.

dispositions to behave in certain self-interested or self-focused ways. Even though Proust does not posit the existence of an independent self, she seems to posit the need for certain self-identifying functions, which she needs to establish continuity of identity of the person. This illustrates the difficulties of putting forward viable accounts of persons without referring to selves. It should be noted that her account is distinct from Parfit's in that her criteria are subjective, not objective, i.e. they are not from a heterophenomenological perspective.<sup>79</sup> This does not mean that they are less scientific or empirical; it is just that her criteria rest on the capacity of the person to interact with their own thinking, and the processes responsible for that. Thus, personhood is dependent on being a subject that can self-reflect and change. To be effective, this requires an acceptance by the subject of sameness of self/person over time. It assumes that 'I' was once one way and now 'I' am or want to be something else. To be a person requires that that person assume their own continuity as the same subjectivity but with a changeable identity. To me, this is putting a subject or self at the core of personhood.

## 2.11 Conclusion

I have argued that the debate on personal identity emphasises particulate identity such as having personal traits and memories. It deals with issues about what kind of person or self one is, not whether or not one is a self or person. There is a distinction between what it is like to be me, as a particular self or person, and what it is like to experience something from a subjective perspective. There is a phenomenal sense of *being something* whenever thoughts or experiences occur that is self-defining. The personal identity debate does not capture the sense of being someone, the subjectivity or ipseity of experience. It leaves it out, hence unexplained.

I argue that persons and selves are not the same. Personhood is very much a social and legal matter, as Locke claimed. It is constituted by personal relations and societal expectations. A person is a public persona which does not completely reflect the inner world of the person. One's self, on the other hand, provides the subjectivity and the sense of oneself as a particular consciousness. One could envisage losing personhood

---

<sup>79</sup> This is Dennett's term for adopting a third person view of events. *Consciousness Explained*, 1991.

but retaining a sense of self. Proust's wantons could fall into that category, as could some primates. Yet, as Proust argues, one cannot be a person without first being a self or at least a subject. Selfhood is necessary if not sufficient for personhood. Similarly, psychological continuity may be a necessary criterion for personhood but not for the self (although some form of conscious continuity might be necessary to experience a sense of oneself).

Nevertheless, if I want to argue that there is such an object as the self, then it should be possible to establish identity criteria for that self, so that it could be identified and re-identified as the same thing. Physical criteria alone will not serve as both selves and persons seem to be more than just physical bodies. Selves (and persons) are not the same as other objects because they can be described using mental predicates as well as physical ones. This gives an additional criterion to identity, although still problematic. At the same time, establishing determinate identity conditions for any object can be problematic in some extreme cases. This is no more nor less true of selves and persons. Just as the destruction of an object does not mean there was no fact of the matter about its identity at some point in time, the same is true of split brain cases and duplicate selves created through teletransportation machines. It may be that in some cases we cannot decide who has survived or who is who after the event. This does not mean that we could not have done this prior to that point. There was a determinate fact of the matter before the cloning or duplication occurred. Having said that, the subjective nature of human experience does mean that a person will be able to make a claim as to their own identity, even in problem cases. This additional criterion is not available to other objects. As such, identity criteria for selves is likely to be different to other objects.

Based on these considerations, I set out the following as necessary and sufficient conditions for something to count as an object and for that object to count as a self.

- Unified boundedness or singularity – Lowe claims that a principle of individuation is really a principle of unity rather than a criterion of identity.<sup>80</sup> An object should be organised in such a way that it is a distinct individual object and there is a fact of the

---

<sup>80</sup> Lowe, p.33, 1998. Gold and water both have identity criteria but they cannot be individuated like ordinary concrete objects.

matter about which object it is. An object should have some internal cohesion that makes it a singular countable thing or its parts should be unified in such a way that the parts work together to form a whole such that the parts cannot be easily separated without loss of identity or function. This means that an object does not need to be a simple substance, but can be complex and still be considered a single object. Most of our ordinary every day objects would fall into this category. For the self to count as an object in this sense it would need to be identifiable either as a singular thing or as a complex but singular-acting thing, a cohesive unit, possibly with some kind of organisational or functional cohesion.

- Synchronic and diachronic identity –Existence, sameness and persistence over time (perdurance or endurance<sup>81</sup>) appear to be essential. If some object can be differentiated from other objects and non-objects, then there must be some kind of identity that that object has that other objects/non-objects lack. This means there is a determinate fact-of-the-matter about which thing in the world some thing is (to qualify as an object) and that it would have its own identity conditions such that it could be differentiated from other things. An object would need some kind of spatio-temporal extension. It should be possible to say that it is here, now or there, then. It should be locatable and identifiable as the same thing through different places and at different times. A further condition of identity is what makes an object a particular kind of thing, what makes it fall under a sortal term.<sup>82</sup>
- Concreteness - Objects can be considered abstract or concrete objects. Abstract objects do not exist in space and time and, as such, are not subject to measurement or change.<sup>83</sup> Concrete objects are generally those objects which have spatio-temporal

---

<sup>81</sup> There are two views of time - the 3D view has tensed objects with no temporal parts; the 4D view has tenseless objects with temporal parts. An object endures (persists through time) if it is wholly existing in any given moment of time. It is not stretched out in time and thus has no temporal parts other than those in the present. An object perdures (persists through time) if it has different temporal parts at different times that it exists but there is no movement through time. See Lowe, p.84-96, 1998 and David Chalmers, 2009 for a discussion of the differences. David Lewis is one of the main proponents of a 4D view of time. See 'Tensed quantifiers' in Zimmerman's *Oxford Studies in Metaphysics: Vol. 1. 2004*.

<sup>82</sup> Lowe says that basic kinds or primitives have only the first condition but not the second. Ie they have determinate identity conditions but no specifiable criterion of identity (it is not possible to say what thing it is). He claims that 'persons' fall under this category. I would want 'selves' to be the same. (1998, p.61)

<sup>83</sup> Again, this is not always the case. Selves are often categorized as 'abstract' to indicate a lack of substance, hence non-real (Dennett) but Reichenbach places complex objects into the abstracta realm.



properties and are objectively detectable or measurable in some way. They are also susceptible to change. If the self were a concrete object, then it should be measurable or detectable in some way, even if this is indirectly as with the evidence for atoms. Interestingly, if the self were concrete, it would be susceptible to change. Yet its supposedly changing character undermines its existence as a re-identifiable object.

- Agency - this relates to an object's causal capacities. For something to exist, it would have causal powers. For the self to exist, it should play a role in the cognitive network such that it has an impact on actions. Ideally, it should be possible for there to be evidence of self-directed action as well as evidence of eliciting self-responses.

One could argue that this last condition is too strong and rules out humans who are cognitively impaired, paralysed, suffering from motor neuron disease or other debilitation which would remove their capacity for action and reaction. I would argue that as long as there was the potential for agency, whether it is realised or not, then it would satisfy this condition. If one were born so deplete of cognition and any voluntary capacity to move, then it is likely there is little potential for personhood or selfhood to develop; there would be no capacity to do. However, there may well be some experience of subjectivity.

- Subjectivity – This is probably the most important for selfhood, inanimate objects cannot specify their own natures or their own identity conditions. Humans can and have this capacity. They can self refer. Thus, there are other individuating mechanisms we can draw on that are not available for other objects, life forms or even some animals. This is the experience of having an internal or mental realm, distinct from the physical (but not separate from it): There must be a subject of experience such that the experiences are not just owned or identified as belonging to x but they are felt by x to be x (selfness quale).

In the following chapters, I set out to show that the self is something rather than nothing and that attempts to negate or eliminate the self leave something about our human psyche unexplained. In the next chapter, I address an alternative conception of the self that emerges out of the Personal Identity debate, the positing of the self as a narrative. This has become a mainstream view in several areas of study that deal with the self – in

particular psychology and some areas of neurophilosophy. I argue against the ontological position that the self is *just* its narrative. I argue that narrative accounts of the self, like the personal identity debate, largely ignore the experiential nature of human existence and are, therefore, inadequate as an account of the self.

## CHAPTER 3

# THE NARRATIVE SELF OR THE SELF-AS-NARRATIVE

I am just a story about a me who is writing a book. When the word ‘I’ appears in this book, it is a convention that both you and I understand, but it does not refer to a persistent, conscious, inner being behind the words. <sup>1</sup>

### 3.1 Introduction: Personal identity and narrative identity

In the last chapter, I argued that the self should not be considered identical to personal identity and that reducing the self to just its identity was not sufficient to address the problem of the self. It left too much of our self-phenomenology unexplained. I also argued that it was problematic to reduce the self to just its properties and that the ontological consequences of doing so applied equally to all objects. While there may sometimes be an issue in providing determinate identity conditions for the self in some circumstances, this problem is not unique to the self and is shared by other objects. Pushed far enough and in certain extreme circumstances, our intuitions can fail to provide a determinate answer to the question, is it still the same object? There will be no fact of the matter either way. However, this does not entail that there never *was* a fact of the matter, nor that objects don’t exist, at least not within a common sense ontology. This also applies to the self.

The idea of the self as a narrative is a natural consequence of the debate about the self and personal identity. The positing of a self-narrative can be found in a wide range of disciplines and has become a widely accepted position within philosophy and other related fields, so much so that it is largely unquestioned. The self-as-narrative, narrative identity, the self-story or autobiographical selves are mainstream concepts that can be found in texts on education, sociology, psychology, philosophy and neuroscience.<sup>2</sup>

---

<sup>1</sup> Susan Blackmore, *The Meme Machine*, Oxford University Press, 1998, p. 230.

<sup>2</sup> see [www.canisius.edu](http://www.canisius.edu) for a website devoted solely to narrative theory and personal identity with several hundred texts referenced from sociology, psychology, neurophysiology and philosophy.

Narrative therapy is a therapeutic tool used to ‘rewrite lives’ that are problematic or damaged by encouraging clients to tell (and retell) their story in a more positive way.<sup>3</sup> Positing the self as a narrative that is constructed during the course of one’s life seems to explain the changing nature of that self, while acknowledging the influence of our social and cultural environment on the identity of that self.

Within philosophical circles, the broad acceptance of narrative theory and the idea of the self-as-narrative can be seen as a response to the difficulties of positing an unchanging, eternal self. First, it appears to remove the need to posit an inner or mental (Cartesian) self. Second, it removes the problem of the ontological status of selves. The self that features within a socially-constructed story about a self has no concrete existence and needs no fixed identity criteria.<sup>4</sup> Third, it legitimates the retention of the term ‘self’ as it still refers to something. Fourth, it supposedly captures both the phenomenology and the heterophenomenology of the self insofar as the narrative creates the impression of singularity, unity and self-agency.<sup>5</sup> This means that the problems raised by trying to ascertain determinate, continuing identity of either a self or a person are no longer relevant and can be bypassed. The narrative provides the cohesion to an otherwise disparate collection of mental items.

However, the advantages of positing a narrative self need to be measured against its feasibility. In this chapter, I want to argue against the ontological assumption that the self is ‘just a narrative,’ that there is nothing more to our selfhood than the telling of a story. I want to demonstrate that the self-narrative is not a sufficient explanation for the way we act, the way we feel, nor the way we experience ourselves as self-conscious agents. The existence of a self-narrative does little to explain or address our self-phenomenology. While it might capture something about how we act in the world, it does not capture our self or our sense of self. I will argue that there is more to the self

---

<sup>3</sup> See Narrative therapy <http://www.narrativetherapycentre.com> - ‘narrative practices collaborate with people in ‘re-authoring’ their lives.’

<sup>4</sup> However, as discussed in Chapter 1, ontological existence is messy and positing the self-as-narrative doesn’t get rid of the problem. A story might still be said to exist but not the characters in the story.

<sup>5</sup> In *Consciousness Explained*, Dennett uses the term ‘heterophenomenology’ to describe the third person perspective or positing things from the outside. It seems to the observer that people are unified, singular, conscious, etc. It is a behaviourist position.

than its narrative, and that there is a self that the narrative is about that is at the heart of the narrative. At the same time, it will be clear that positing a narrative *on its own* does not do the work it needs to.

It should be noted that not all narrative accounts of the self claim that there is no self or that the subsequent narrative self is not real. Ontological positions on the self-narrative vary or are not always elucidated. Such proponents may consider the resulting narrative real, a real abstraction or an existing entity, depending on their own ontologies. While this further demonstrates the ontological confusion in relation to what the self is, this is not my concern in this chapter. Narrative accounts are also not uniform in how they depict the relationship between the narrative and the self. Some treat the narrative as an expression of the self, rather than the creator of the self. Proponents of a minimal or core self, may posit the narrative or autobiographical self as an extension of the basic self.<sup>6</sup> It is outside the scope of this chapter to address all the variations.<sup>7</sup> What I am most concerned about is the proposition that the narrative, given there is such a thing, is *all there is* to the self, and that the narrative fully captures our self-hood. It is this point that I take issue with here.

After a brief introduction to the role of narrative in self-construction, I will present the work of three proponents of a narrative self: Dennett, Velleman and Schechtmann. They represent the strongest and most influential accounts in the philosophic literature and they differ from each other in interesting ways. I will start with an exposition of each writer followed by a critique. During this discussion, the inherent weaknesses in the narrative model of the self will become apparent. I go on to raise more general issues in relation to the concept of narrativity and argue that it is by no means obvious that our self-identity is narrative in form and that positing the narrative creates problems for our

---

<sup>6</sup> See P. Boyer, P Robbins and A Jack, 'Varieties of self-systems worth having,' *Consciousness and Cognition*, Vol 14, 2005, pp 647-660, for an explication of these. Gallagher claims recent literature on the self can largely be represented by two main approaches – those that focus on the 'minimal' sense of self and those that focus on the narrative self (2000). I would argue there is more diversity than that as 'core' self and 'minimal' self can be conceived differently. I refer to this point in Chapter 6.

<sup>7</sup> Marya Schechtmann gives an interesting overview of various narrative positions and their problems in 'The narrative self', *The Oxford Handbook of the Self*, edited by Sean Gallagher, OUP, 2011, pp. 397-416.

identity. I use Strawson to flesh this out. Zahavi highlights the limitations of the narrative to fully capture the self-phenomenology.

In particular, Zahavi argues that the use of the narrative self (or its absence) as an explanation of all ills puts it at risk of becoming explanatorily empty. I will conclude that the self cannot be adequately captured by using just a narrative construction for the same reasons the self is not explicable by nor reducible to just personal identity. In arguing my point, there is likely to be some overlap between my critique here and my critique of personal identity in Chapter 2.

### **3.2 Positing the self as a narrative and self-narrating as normative**

In ‘The Neurology of Narrative,’ Young and Saver discuss the importance of telling stories or ‘narratives’ to help us understand ourselves and the world in which we live.<sup>8</sup> Telling stories, they say, is natural and no human culture has existed without stories. They go on to suggest that the capacity to narrate a story, particularly one’s own story, is a necessary feature of human development. So essential a part of human development is it that lacking the capacity to tell stories about oneself indicates some kind of pathology. On the other hand, constructing a narrative about one’s life is not only normative but it makes self-understanding possible through the telling of the story. They go on to claim that the narrative constructs who one is. Referring to Schank’s work on understanding intelligence, they state that ‘[t]o be without stories means...to be without memories, which means something like being without a self.’<sup>9</sup> For narrativists like Young and Saver, the story IS the self because it captures who one is, i.e. one’s personal identity. There is nothing more to the self than this identity, the story we tell about ourselves. And like any autobiography, it is an interpretation of events and actions, carefully selected and edited to present oneself in a particular light.

---

<sup>8</sup> Kay Young and Jeffrey Saver, ‘The Neurology of Narrative’, *Substance* 94/95, 2001 pp. 72-84.

<sup>9</sup> Young and Saver, 2001, p. 74. While not directly mentioned, this sentiment assumes both that one’s identity is memory-based and that the self is akin to identity.

This view is far from unique. Paul Ricoeur claims that ‘narrative identity’ is essential to our very existence.<sup>10</sup> It is through the narrative that our actions take on meaning, that they make sense, that they become digging a hole or building a house rather than merely bending and lifting. According to Ricoeur, action, to be defined as action, needs purpose and the purpose comes from where that action fits in progressing the narrative, in carrying it forward to its next stage. Without our narrative in which we feature as the main character, or even the plot itself, we would have no reason to act, as opposed to merely move. Our lives would lack meaning.

Do we not consider human lives to be more readable when they have been interpreted in terms of the stories that people tell about them? And are not these life stories in turn made more intelligible when the narrative modes of plot – borrowed from history or from fiction – are applied to them? ...self-understanding is an interpretation; interpretation of the self, in turn, finds in the narrative...a privileged form of mediation; the latter borrows from history as well as from fiction, making a life story a fictional history or, if one prefers, a historical fiction.<sup>11</sup>

He claims that the narrative is both an interpretation of the past and a projection into the future. As such it is constantly unfolding as new events and situations occur in one's environment. Ricoeur claims that the narrative makes it possible for others to understand our actions, as well as ourselves. Narrating one's self story is analogous to using the form of any standard autobiography. A backward referring historical account of one's life and the events that have shaped it is created, including justification and interpretation of one's actions, to form a coherent and convincing tale. As Ricoeur says, we understand people, their personalities, their motivations, their actions *because* there is a story to be told about who they are and why they are. Without this context, people's actions would appear meaningless.

While it may be the case that we do tell stories about ourselves and that this makes understanding each other easier, this by no means entails that this is anything more than an expression of our capacity to talk about ourselves as members of a linguistic

---

<sup>10</sup> Paul Ricoeur is taken as the originator of the concept of ‘narrative self.’ See *Oneself as Another*, trans. Kathleen Blamey, University of Chicago Press, Chicago, 1992. Galen Strawson credits Alisdair McIntyre in *After Virtue*, 1981, with being the originator of the idea. See ‘Against narrativity,’ *Ratio*, Vol. 17, 2004.

<sup>11</sup> Paul Ricoeur, 1992, p.114, footnote 1.

community. The narrative mirrors who we think we are. However, those that posit a narrative self of some kind go beyond the psychological benefits of story-telling and claim that it is the having of a narrative that constitutes who we are as selves and persons. They claim that the narrative is what we mean by having a self. You are your story. As Blackmore makes explicit in *The Meme Machine*, there is either a real, eternal non-physical self or there is an illusory self-story spun by the brain. She claims logic and evidence rest on the side of the illusory self-story.

I am just a story about a me who is writing a book. When the word 'I' appears in this book, it is a convention that both you and I understand, but it does not refer to a persistent, conscious, inner being behind the words.<sup>12</sup>

Blackmore, like most narrativists, is making an ontological claim about the self. Either there is a Cartesian (non-physical) self or there is no self. Her conclusion is that the self is not an existent thing in its own right but a product or fantasy of the self-story.

The story might exist (although several versions of this position imply that the story is also fictitious), but there is no real self that the story is about. The (Cartesian) self only exists as a character in the narrative. Its nature and capacities are invented, producing the illusion of a singular, persistent, agential self.

In the next few sections, I discuss three influential accounts of the narrative self that put forward similar claims. Schechtmann's emerges directly from the personal identity debate, whereas Velleman acknowledges Dennett's influence (as does Blackmore in *The Meme Machine* above). Dennett's account has been influential beyond the field of philosophy. I will begin with an exposition of each account and follow this with a critique. I claim that none of these accounts are viable, but for different reasons. Amongst other things, both Schechtmann and Velleman rely on some form of subjectivity or inner self that is not accounted for within the narrative, whereas Dennett denies phenomenal consciousness altogether. All of them suffer from the short-comings of positing just a narrative to deal with the phenomenon of selfhood. I finish with a general critique of the self-as-narrative by referring to both Strawson and Zahavi's work.

---

<sup>12</sup> Blackmore, 1998, p. 230.



### 3.3 Schechtmann: the self, personal identity and the construction of persons

In *The Constitution of Selves*, Schechtmann puts forward an account of the role of narrative in self-construction, known as the Narrative Self Construction View or NSCV.<sup>13</sup> Although she uses the term ‘self,’ she views the term ‘self’ as synonymous with ‘personal identity.’ For Schechtmann, persons are distinct from selves, and the term ‘self’ is just another way of talking about our (personal) identity. This means that self-construction is really identity construction. Her principal objective in *The Constitution of Selves* is to argue against reductive accounts of personal identity like Parfit’s and to put forward what she says is a neo-Lockean account of persons that addresses the problems of psychological continuity.<sup>14</sup> She claims that most accounts of personal identity conflate one set of problems with another; namely issues about re-identification (identifying the same ‘you’) with issues about characterization (how ‘you’ is constituted). The key elements of what constitute personhood are mistakenly taken to be matters that re-identification must account for, leading to accounts that are problematic and which bear no resemblance to our lived experience of being persons, such as Parfit’s impersonal mental events. (Ricouer raised a similar issue in his critique of Descartes and Parfit, mentioned in Chapter 2.)

According to Schechtmann, there are four features necessary for personhood – survival, self-interested concern, moral responsibility and compensation. In this, she reflects Locke’s conception of a person as ‘forensic,’ necessitating social, legal and moral responsibility.<sup>15</sup> Each of these features is linked to one’s identity because they rely on the continuity of the person over time. For example, compensation is only due to this person on the grounds that this person is continuous with that same past person. Nevertheless, Schechtmann argues that these are issues of characterization, not re-identification. She puts forward a ‘narrative self-constitution view,’ the claim that the

---

<sup>13</sup> Marya Schechtmann, *The Constitution of Selves*, Ithaca, Cornell University Press, 1996.

<sup>14</sup> Andrew Lane, ‘The Narrative Self-Constitution View: Why Marya Schechtman Cannot Require it for Personhood,’ *Macalester Journal of Philosophy*, Vol. 20, Issue 1, 2012.

<sup>15</sup> I referred to this aspect of Locke’s work in Chapter 2. John Locke, Bk II, Chap. Xxvii, s26 , *An Essay Concerning Human Understanding*.

self is constituted by its own narrative, as a means of addressing these features. Schechtmann argues that the actions through which the four features are realised are dependent on a self-concept. The narrative provides the self-concept that provides a conception of oneself as a particular kind of person. Or to put it another way, an identity-constituting self-conception is one provided by the self-narrative. In Schechtmann's account, if there is no sense of (one's) self then there is no person; if there is no self-concept then there can be no means to achieve personhood. A person,

...creates her identity [only] by forming an autobiographical narrative – a story of her life. She must be in possession of a full and explicit narrative to develop fully as a person.<sup>16</sup>

Schechtmann uses the self-narrative concept to address problems of continuity and determinate identity criteria. Continuity, she claims, is inherent in the narrative account. She defends what she sees as Locke's account of personal identity by claiming it relied on more than simple memory. For Locke, it is the *conscious* engaging with experience that makes that experience one's own.<sup>17</sup> Her own version of personal identity relies on an individual's 'subjective relation to her actions and experience' in that the contents of the narrative are what the person has appropriated as her own or what she takes to be important or self-constituting.<sup>18</sup> According to Schechtmann, the self-narrative account rests on the assumption that only certain kinds of sentient creatures are persons and that the difference between persons and non-persons lies in how one organises one's experiences. For persons, this is done by creating a self-narrative, a process of weaving events in one's life into a coherent story. The narrative is self-constituting because it defines who one is and what matters. It provides the theme to one's personal identity. In this way, the narrative represents the person's inner life as much as her outer, social life. We see reflections of this role of the narrative in Ricoeur. Schechtmann takes the narrative analogy literally. As with Ricoeur's account, it really is like a literary genre. She claims that the narrative should take the form of a traditional linear autobiography with a past, a present and a possible future. It must follow the logic of a story. According

---

<sup>16</sup> Schechtmann, 1996, p. 93.

<sup>17</sup> See Chapter 2. John Locke, Bk II, Chap. Xxvii, s 9

<sup>18</sup> Schechtmann, 1996, p.95.

to Schechtmann, one should be able to expound one's life story, or self-narrative, in the same way one can retell other stories - in a story-like manner. Schechtmann places constraints on the contents of the self-story. Not any story can be told about oneself. It must be articulable, it must be internally coherent to some extent and it must conform to known events in the person's life. This limits the extent to which it can be fictitious; there are reality constraints and these serve to provide one's identity with a more stable character. The self-narrative should bear a close resemblance to events that have actually occurred in one's life. If one behaves in what appears to be an uncharacteristic way, there should be some plausible explanation that fits with what people know of you, or there should be an acknowledgement that the behaviour was an anomaly. If there is too little coherence, then this could undermine one's status as a person. Schechtmann acknowledges that some narratives will be more coherent than others and that there is debate about at what point one's personhood would be threatened. However, she claims that for us to satisfy the four features, we need a reasonably coherent self-constituting narrative. Beings with a radically different narrative or no narrative at all will not count as persons. Under this aegis, someone with dementia, advanced Alzheimer's or some psychoses would lose their status as persons. Some other severely (mentally) disabled individuals may never achieve personhood. According to Lane, and as a critique of Schechtmann, humans from radically different cultures and beliefs may also be in danger of not qualifying.<sup>19</sup>

Against Parfit, Schechtmann argues that continuity does not rely on just memory. Past experiences can leave effects and affective traces in the brain somewhere, even if the actual memory is gone or we are unaware of the effects that past events have had on our present identity. The past shapes the present and the future self, even if we have lost the memories. As she claims in 'Personal identity and the Past,'

There is an important difference between an experience that is mine because I experienced it in the past but have now forgotten it entirely, and one that is mine because I have repressed it and am still suffering the symptoms of that

---

<sup>19</sup> Lane raises this point as a critique of Schechtmann's account, 'The Narrative Self-Constitution View: Why Marya Schechtman Cannot Require it for Personhood,' 2012.

repression, and there is no clear way to capture this difference in the psychological continuity theory.<sup>20</sup>

This means the past can be present in the present self, even if we have no *conscious* memory of a particular event.<sup>21</sup> Likewise, the past can set conditions or leave environmental traces which influence future acts. Assumptions of wealth, health and social networks, or their lack, will influence lifestyle choices and future intentions.<sup>22</sup> The continuity of our environs provides another kind of temporal continuity to personal identity and the link between past, present and future stages. Schechtmann wants to demonstrate that there may not be enough connectedness or an obvious causal chain linking one event to another to satisfy Parfit's Relation R yet this doesn't prevent those things having an impact on who we are and how we act. Past events, emotional reactions, beliefs may no longer be memories or may not have become memories, yet they can still play a role in one's current psychology. It also acknowledges that persons are situated in a socio-cultural niche. The narrative will express a particular content that assumes and reflects the impact of past and present on a future self or person.<sup>23</sup> It will reflect a character.

Schechtmann assumes that the self of the narrative is not an independently existing entity. She also assumes that there is nothing in our neurological make-up that could be classified as a self or self-like structure.<sup>24</sup> The self is purely a concept or an idea which provides coherence to what would otherwise be Hume's bundle of experiences. For her, the self and all its phenomenology is fully captured by reference to a person's identity via the narrative. The narrative is not reducible to a particular cognitive structure.

---

<sup>20</sup> Marva Schechtmann, 'Personal Identity and the Past' in *Philosophy, Psychiatry, & Psychology*, Vol. 12, No.1, March 2005, p 16.

<sup>21</sup> Kathryn Wilkes raised a similar concern in relation to Strawson's purely mental self which I address in Chapter 7. See 'Know Thyself,' in *Models of the Self*, editors S Gallagher and J Shear, Imprint Academic, U.K., 1999, pp. 25-38.

<sup>22</sup> Schechtmann, 2005, pp. 9-22.

<sup>23</sup> See Schechtmann, 2005, pp 9-22.

<sup>24</sup> Like Parfit, Schechtmann does not talk about brain processes. She is not really interested in the neuroscience or how it happens in the brain. Her focus is the debate on personal identity and self-constitution.

Once we see ourselves in the narrative, we have a self-concept. Unlike Hume and Parfit and most other narrativists, she does acknowledge subjectivity and claims this aspect puts Locke's psychological continuity criterion beyond just simple memory. Pace Parfit, this could make a person a conscious entity and at least means there is an acknowledgement that there is some subjective phenomenology going on. However, that subjectivity does not seem to play any kind of role in the self-narrative other than as a mechanism for providing a link between mental states. It does not entail a subject. It seems that its role in Schechtmann's account is to bolster claims of temporal continuity.

Schechtmann does provide some way out of the Lockean problem of psychological continuity and reliance on memory. She does offer a mechanism other than memory that would link past selves with present selves, based on the assumption that selves are either temporally identifiable discrete units or synonymous with personal identity. However, she gives no credence to a self that may be an independently-existing entity or that is something in addition to the narrative. The existence of a self is dependent on the presence of a self-concept produced through the narrative, and this makes it the same as personal identity and, ultimately, reducible to an idea. This sort of self lacks an obvious causal role and would be bereft of any kind of agency. The extent of its singularity or unity would be dependent on the quality of the narrative, nothing more.

### **3.4 Velleman and multiple selves**

Velleman's is an interesting contrast because his conception of the narrative self is a direct response to Dennett's account. He claims to be sympathetic to Schechtmann's narrative self-constitution view and to have only minimal disagreement with Dennett, which I discuss below. He has, however, written extensively about the self in other ways in an attempt to capture what he claims the self is. Close reading of his papers reveal a profusion of ideas about the self, of which the narrative appears to be just one. The difficulty with Velleman's account is trying to work out how his collection of other supposed 'selves' fit together.

As a starting point, Velleman does not think the term ‘self’ refers to a singular item or thing.<sup>25</sup> In the introduction to his collection of essays, Velleman makes it clear that he does not think the self is an entity as is commonly conceived. For Velleman, the self ‘expresses a “reflective guise” under which parts of oneself are presented to [one’s] own mind.’<sup>26</sup> Like Sloman and Kenny, he claims the term ‘self’ merely captures a linguistic expression; this is how we talk about ourselves from the inside.<sup>27</sup> This does not mean there is a self in addition to the contents. It just means I remember *me* going to the shop, or I sat *myself* down rather than someone else making me do it. So, while assumptions of autonomy and ownership are intrinsic to the meaning of the term ‘self,’ the term ‘self’ actually refers to some of the many reflexive capacities of a person, rather than to a separate entity. He says the term ‘self’ has various uses in various contexts.<sup>28</sup> According to Velleman, what this indicates is that there are various aspects of the person with various capacities to think ‘I’ thoughts and to own certain actions or to engage in other mental activities. Being reflexive means that the object or subject of a thought turns out to be either ‘me’ or ‘I.’ The self is a mode of first-person reference and the experience of a first-person perspective.<sup>29</sup> Velleman says that when we refer to ourselves as ‘I’ or ‘me,’ this is not singling out something unique. We are actually referring to a person’s capacity to do something mentally (although it is by no means clear what Velleman means by this).

Velleman claims that philosophical mistakes are made in regards to discussions about the self because 1) it is assumed it is a single entity and 2) it is assumed that it is the same entity through different modes. Velleman wants to show that the one concept can’t deal with the range of so-called ‘self’ events and that it is best to talk about multiple

---

<sup>25</sup> See J. David Velleman, *Self to Self: selected essays*, Cambridge University Press, New York, 2006, and Velleman, ‘So it goes,’ *The Amherst Lecture in Philosophy 1*, 2006, pp 1-23, <http://www.amherstlecture.org/velleman2006/>.

<sup>26</sup> Velleman,, ‘Introduction,’ *Self to Self*, 2006, p.1.

<sup>27</sup> Aaron Sloman, ‘*The Self*’- a bogus concept, accessed November 2010 <http://www.cs.bham.ac.uk/research/projects/cogaff/misc/the-self.html>, and A. Kenny, *The Self*, Marquette, Marquette University Press, 1998.

<sup>28</sup> Velleman, ‘Identification and Identity’, *Self to Self*, 2006, p 354.

<sup>29</sup> Velleman refers to this capacity in several essays in his volume of essays. See the chapters ‘Self to self,’ and ‘Identification and identity,’ *Self to Self*, 2006, pp. 192-194 and pp. 354-356.

‘selves’ to represent different aspects of the same person. He claims that whenever the self is referred to, some particular reflexive state is being identified, just as when we talk of the subject there has to be a subjective state that the subject is subject to or of.

What Velleman does to address this is to distinguish three different reflexive guises which he goes on to call three distinct ‘selves’ – self-image, self-sameness and autonomous agency.<sup>30</sup> The self-image is akin to Dennett’s self-representation ‘blip.’<sup>31</sup> It picks out the external manifestation of who you are in terms of autobiographical details and by tracking the physical body that is you. This is your external or objective identity, what individuates you (or your body) from others. Velleman claims that this embodies one’s sense of self. By this, he does not mean an internal subjective experience of selfness or self-phenomenology (as I and others use the phrase) but who you are as a distinct person – me as opposed to you. Velleman claims that this guise is *non-reflexive* because you identify yourself the same way you identify others, using objective means. At the same time, you identify yourself as the subject of the self-identifying information (it’s about me). Thus, one’s self-image will constitute who you are as a certain kind of person. It *becomes* reflexive because one can reflect on one’s self image and (try to) conform one’s actions to one’s image and vice versa.

For his second guise, Velleman draws a distinction between his self-image of personal identity and the self-sameness guise of psychological or subjective continuity. Velleman claims that the self that identifies itself through engagement with past selves via first-personal access to memory, is distinct from its identity as a particular kind of person. Rather confusingly, Velleman refers to this present subjectivity as a self and past embodiments of subjectivity as past selves, creating a plethora of temporal selves. It is the self-sameness guise that provides the sense of continuity, through engagement with one’s internal states like memory and experience from one’s point of view. It provides the ‘I’ of identification of these states. In most other accounts, these aspects would be considered part of one’s personal identity, as it is access to one’s memories that

---

<sup>30</sup> Velleman explains these in ‘Introduction’ in *Self to Self*. His choice of terminology is confusing as it conflicts with his later claims of multiple temporal selves.

<sup>31</sup> Dennett, p. 429, 1991.

contribute towards who you are and claim to be. Such a faculty would then contribute to the self-narrative. However, this is where Velleman differs from other accounts. He claims that personal identity is distinct from the self-sameness self, although both provide identity criteria for persons, and that these both are distinct again from the narrative self. The narrative is the third guise and, for Velleman, is the provider of autonomous agency. According to Velleman, that part of the person that is presented to causal reasoning is the thing/self that is doing the reasoning or attempting to understand, i.e. it represents the faculty of causal understanding. The agent's faculty of causal understanding is seen as the self that is responsible for the actions the person performs.<sup>32</sup> The action is autonomous if performed for a reason, and that reason is what makes that action understandable to the agent in terms of belief/desire psychology. The story or narrative that is told about the action provides the reasons for action and that in turn becomes the cause of one's future actions. For Velleman, the story also gives one the sense of ownership and control over one's actions. Action requires reason and the narrative provides the rationale through the demands of coherence. According to Velleman, one is more likely to perform an action because it fits one's narrative self-conception than because it is necessarily the best action to take. Velleman views the narrative as self-fulfilling. You literally become the self of the narrative and, as such, can claim that there is (now) a real self. In this respect, he differs from Dennett.

If a self-narrator works in both directions then the self he invents is not just an idle fiction, a useful abstraction for interpreting his behavior. It...is a determinant of the very behavior that it's useful for interpreting.<sup>33</sup>

For Velleman, the narrative module must ensure that the narrative is internally coherent and that it coheres with the 'real' life story of the narrator. This means there is a two-way correspondence between the life as acted out and the story, with each adjusting to the other and, in effect, causally determining each other. Velleman actually claims that the capacity to determine action by stipulating it in one's narrative is akin to free will.<sup>34</sup> Unless one sub-routine is much stronger than others, the 'human self-narrator' can

---

<sup>32</sup> Velleman, 'Introduction,' *Self to Self*, 2006, pp. 7-9.

<sup>33</sup> Velleman, *Self to Self*, 2006, p. 212.

<sup>34</sup> Velleman, *Self to Self*, 2006, p. 212.



choose between subroutines and act them out.<sup>35</sup> The decider will be narrative coherence rather than reasons to perform one act over another. Rational justification emerges from narrative continuity not vice versa. Thus, ‘I’ make a choice because it fits with who I am as part of my narrative. In this sense, the narrator and the narrative are interdependent.

Velleman claims several different things about the self that make it hard to get a clear picture about what his ‘selves’ actually are in his schema. Velleman is very definite in the introduction to his account about two things. One, there is no self and what we call the self is actually a diverse set of capacities. Two, if the term ‘self’ refers at all it is to the narrative. There is nothing else to which it refers. However, he wants to claim some reality to the constructed self of the narrative. If this self does come into existence through the process he claims, then there is a something that is that self and there is a something to which the term refers, even if it is only the self within the narrative. That self, so he argues, is capable of autonomous action. It is causal. In fact, his only dispute with Dennett is about the consequent reality of the narrative self. As he says in relation to Dennett’s hypothetical story-telling robot, Gilbert,

Now that the robot has a central controller that makes decisions for reasons, he has a self, and so his story has become true... [But] Gilbert is not the name of a self; it’s the name of a unified agent who has a self, in the form of an inner locus of control.<sup>36</sup>

So the narrative self, hence our everyday self, has taken on a reality. Not only that, but this ‘inner narrator as a unified self’ has ‘agential unity’ in virtue of being self-governed.<sup>37</sup> So this is a real, unified, autonomous self, reminiscent of a Cartesian self.

Earlier, Velleman mentioned three ‘selves’ or modes as three distinct capacities of persons that were not able to be captured within one concept of the self. Yet here (above) he appears to actually advocate a unified concept through his explication of the narrative self. Velleman has adopted Dennett’s conception of the self as a *centre* of narrative gravity, which means the narrative acts as a unifying and controlling

---

<sup>35</sup> Velleman, *Self to Self* p. 213.

<sup>36</sup> Velleman, *Self to Self*, 2006, pp. 220-221.

<sup>37</sup> Velleman, *Self to Self*, 2006, p 223.

mechanism. It ties the other guises together. His only disagreement with Dennett is in how fictional the narrative is. He claims it is not fictional as in *unreal* but is fictive in that it is made up.<sup>38</sup> He goes on to claim that we actually become the selves we invent. In fact, the autobiography becomes self-constituting. Thus the fictional narrator really does become the self or the central controller, as the story will dictate action just as much as action determines story. Consistency and coherence will lead to predictable behaviours in some circumstances and telling a story about what you intend to do will often mean that you do those things you said, *because* you said them. Velleman cites experimental evidence that demonstrates how we conform our lives to our stories; if we believe we are angry, we will act angrily; if we say we will do something, we are more than likely to end up doing it. Thus, it seems that the self-sameness and self-image guises will all fall under the self-narrative guise as parts of the one narrative self.

As a further complication, though, Velleman seems to hold the view that there is something else that can be called the self, represented by the self-sameness guise. In ‘Self to self’ and ‘Identification and identity,’ he refers to the Lockean idea that continuity over time is provided by being able to experience past memories as one’s own. Velleman says this is the capacity to engage with a memory from a first-person perspective and with first-person access. He uses the example of imagining being Napoleon to flesh this out.<sup>39</sup> If you say, ‘I am imagining I am Napoleon,’ the two I’s do not refer to the same thing. He claims that the term ‘self’ has two related meanings.<sup>40</sup> The first is a ‘metaphysical relation that holds between persons’ over time, while the other is a ‘psychological relation that holds between subjects’ who share first-person perspectives.<sup>41</sup> For Velleman, this means I can engage with a past event via memory and claim it is ‘me,’ without claiming I am the same person. My identity will not be identical to that prior ‘me’ but I will share the same subjectivity because I have the capacity to

---

<sup>38</sup> In *Self to self*, Velleman distinguishes between ‘fiction,’ as an example of something unreal like a unicorn, and ‘fictive’ which uses the story-telling genre to effect. See p.221 in ‘Self as narrator’ for an explication which contrasts ‘fictive’ with Dennett’s fictional narrative.

<sup>39</sup> Velleman, ‘Self to self,’ *Self to Self*, 2006.

<sup>40</sup> See Chapters 1 and 2. I have termed this as the difference between ‘who I am’ and ‘that I am,’ the distinction between identity and existence; see Ricouer’s ipse/idem distinction.

<sup>41</sup> There are parallels between this distinction and Strawson’s distinction between ‘I’ and ‘I\*.’ Velleman, *Self to Self*, 2006, p.192-3.

engage with it from a first-person perspective. I will be using the first-person pronoun correctly. Velleman then says that selfhood is not the same as personhood. If I can genuinely engage with Napoleon's memory then he is a past self of mine and I can refer to him as 'me.' But I am not the same person as Napoleon; I am not Napoleon. So Velleman appears to make another distinction about selves. A self in this context appears to be the *subjective* experiencer of first-person events in contra-distinction to the person. The self is the same as the subject. In addition, he assumes there is no genuine continuity of the subject, only a psychological relation between different episodic subjects. This means there are multiple selves in any one system because each past self/subject is a self 'time-slice.' When Dennett posits the narrative, it is to weave a continual unifying tale over the whole life.<sup>42</sup> In contrast, self-narration for Velleman is local, consisting of 'small, disconnected stories about ourselves - short episodes that do not get incorporated into our life-stories.'<sup>43</sup>

Velleman's account of the self seems to confirm the difficulty of not talking about a self or of not positing some kind of inner entity, even when we want to avoid positing such an entity. Although Velleman starts by dismissing the self altogether as just a self-referential term, he proceeds to posit multiple selves, both temporal and present that appear to be much more than just guises or terms of reference.

### **3.5 Dennett and the centre of narrative gravity**

In 1991 Dennett wrote *Consciousness Explained* in which he attempted to explain both what consciousness, the mind and the self were and how they could come into being through natural selection and within a purely physical system.<sup>44</sup> He also promised to do this by not 'feigning anaesthesia' or ignoring his own phenomenological experiences.<sup>45</sup> His approach was self-consciously anti-Cartesian, claiming to finally eradicate the need to posit an internal witness or any form of autonomous agent. According to Dennett, the self of the Cartesian non-physical, self-conscious agent could be fully explained by

---

<sup>42</sup> See Schechtmann's discussion, pp 408-416, in 'The narrative self,' 2011.

<sup>43</sup> Velleman, *Self to Self*, 2006, p. 222.

<sup>44</sup> Daniel Dennett, *Consciousness Explained*, Little Brown, Boston, 1991.

<sup>45</sup> Daniel Dennett, 1991, p. 40.

positing the existence of a self-narrative. He argued that the cognitive system, through the operations of a virtual 'meme machine,' spins a tale about its own exploits which make it appear as though there is an agent or a self inside. In reality, there is no self in the way we think there is. There is just a story about a self and the story or narrative is all there is to the self and its phenomenology. According to Dennett, the self is rather like a Centre of (Narrative) Gravity. As such, it has no concrete reality but can be considered a 'real abstraction.'<sup>46</sup> Dennett claims the (narrative) self really does exist at the level of social discourse but it has no underlying cognitive hardware that is representative of it. Hence, it has no concrete embodiment as such. Because the cognitive architecture is transformed through the invasion of social and cultural information via the meme machine, the self cannot be explained by adverting to biology. The new software radically changes the original hardware of the brain to create a mind, a self-narrative and the myth of consciousness.

Dennett's account of the self, mind and consciousness is counter-intuitive and has been strongly criticised by many.<sup>47</sup> Nevertheless, his account of the self as a centre of narrative gravity has been highly influential and is referred to across a range of disciplines. Dennett's concept of the narrative self is probably the most thorough in that he describes the mechanisms he believes are responsible for its creation. His account of meme invasion (alluded to in the introduction and Blackmore's book), is intended to remove the need to posit an agent of any kind. If the agent is no longer necessary for explanatory purposes, then we no longer need to worry about what it might look like and how a brain produces one. In Dennett's model, the memes become the agent by providing the mental content and the constraints on action. His rationale for positing an impersonal agent is to remove any form of closet dualism whereby the self is present as some form of hidden Cartesian inner witness.

Adverting to neurological research and computational modeling, Dennett claims there is no evidence of a singular self or internal witness inside the cognitive architecture of the

---

<sup>46</sup> Dennett, 'The Self as a Center of Narrative Gravity,' in F. Kesses, P. Cole and D. Johnson (eds.), *Self and Consciousness: Multiple Perspectives*, Erlbaum, Hillsdale, N.J., 1992.

<sup>47</sup> See for example 'Symposium: Daniel C. Dennett's Consciousness Explained,' in *Inquiry*, 1992, Vol. 36 and Bo Dahlbom (ed.), *Dennett and His Critics: Demystifying Mind*, Blackwell, Oxford, 1993.

brain. As a consequence, he offers a reductive account of the self along the lines of Hume and Parfit. However, he does not want to claim that the self is a non-existent thing because he has promised to not ignore our phenomenology. He claims the self does exist but only as an abstraction, albeit one he claims is real.<sup>48</sup> This self, he claims, is produced via the narrative-writing mechanisms of the brain.

As with many narrative accounts, there is no internal reality to the self; the self it describes is not intrinsically singular or unified. The reason we think that the self is singular is because we construct a narrative about ourselves or, more accurately, a narrative is spun by the human cognitive system which creates an impression of a singular self that is also the agent of its actions. He states that:

Our tales are spun, but for the most part we don't spin them; they spin us. Our human consciousness, and our narrative selfhood, is their product, not their source. These strings or streams of narrative issue forth *as if* from a single source – not just in the obvious physical sense of flowing from just one mouth, or one pencil or pen, but in a more subtle sense: their effect on any audience is to encourage them to posit a *center of narrative gravity*.<sup>49</sup>

The main difference with Dennett's account and those of others, in particular Velleman and Schechtmann, is that he is not just interested in narrative identity as a form of personal identity. His focus is not on providing determinate identity criteria for persons or selves. Dennett is interested in explaining how a physical system can, seemingly, produce such odd phenomena as minds, consciousness and selves. In the course of his explanation, he dismantles the common conceptions of the mind and consciousness, as well as the self.<sup>50</sup> To be successful, he has to put forward an account of how these capacities are generated in the human system. The self becomes a product of the system, as does consciousness and the mind.

In itself, this would not be problematic. If we have or are selves, if we have or are minds, and if we are conscious beings then these things must be the products of the kind of cognitive architecture we have. However, Dennett argues that the mind,

---

<sup>48</sup> Dennett discusses this most fully in 'Real patterns,' *The Journal of Philosophy*, 1991, 87, pp. 27-51.

<sup>49</sup> Dennett, 1991, p. 418.

<sup>50</sup> Dennett, 1991, pp. 253-284,

consciousness and the self are not ‘givens;’ it is not inevitable that they exist. They are the product of dramatic changes to the original architecture of the brain through the invasion of ‘memes.’ Memes are socio-cultural informational units that invade the brain predominantly by using the medium of language. The resulting software or ‘meme machine’ creates the illusion of mind, self and consciousness. This removes any causal power or agency from the self. The resultant self-phenomenology is what Dennett calls a ‘user-illusion.’<sup>51</sup> It is a product of the infestation of memes, both because the self-concept is a meme and because the ‘Joycean’ meme machine creates a virtual, singular self. The meme machine transforms the original hardware of the brain in much the same way a chess program transforms a desktop computer from a word-processor into a chess playing machine. Because memes are socio-cultural units of information, the resultant phenomenology will also be socially and cultural specific. At the same time, the only reality accruing to our accounts of ourselves is what we say is the case. There will be no fact of the matter about whether one’s narrative is really true or not. If I say I was conscious of x at t<sub>1</sub>, then I was. What I say is the case will depend on which draft of the story makes it into words. Hence, I learn about myself the way other people do, by listening to what I say and watching what I do.<sup>52</sup> The rest, the supposed inner mental life, is illusion.

Unlike Schechtmann and to some extent, Velleman, Dennett argues that the self-narrative is largely fictitious and that the items in that narrative are illusions. Dennett wants to demonstrate that claims we make about ourselves are not a guide to what we are really like or what it is we really experience. Contra Descartes, there is no self behind the self-phenomenology whose existence needs to be explained. Self-consciousness itself is an illusion. We do not really ‘experience a sense of self’ anymore than we experience pink rings or pains. We just claim that we do because it seems like we do based on the contents of our minds.<sup>53</sup> These ‘experiences’ are examples of first and second order judgments. We are not phenomenally conscious but we can make statements about what we think we experience when probed. At what point we report on

---

<sup>51</sup> Dennett, 1991, pp.275-80.

<sup>52</sup> Dennett, 1991, pp. 275-80.

<sup>53</sup> Dennett, 1991, p. 373.

the contents of our minds will determine what we claim to be conscious of at that time. But, or so Dennett claims, ‘...it does not follow from this...that *there really is* phenomenology. It only seems like there is. There is no way things *really* are.’<sup>54</sup> A moment later and the story would have been different.

According to Dennett, we cannot rely on what we think is going on in our mind because, more often than not, we will be mistaken. In fact, we make up stories or confabulate. He says we don’t know what we will do or say until we do or say it.<sup>55</sup> Thus, introspective knowledge claims are unreliable. The only ‘objective’ criteria will be from the third-person, heterophenomenological perspective. Dennett claims we learn about ourselves in this ‘third-person’ way. If we want to know which thing we are in the world, we ‘do something and “look” to see what “moves”’ (Dennett’s scare quotes).<sup>56</sup> If we want to know what we think, we listen to our stories. According to Dennett, we can’t know anymore than anyone else what is going on in our mind, what we really think or believe or what something is really like from our perspective. Such reports will be theory-dependent, fictional and based on the illusory contents of the self-narrative, some of which will include claims about our ‘supposed’ phenomenal experiences of things.

Dennett’s position has been picked up by others, notably Blackmore (and Claxton, cited in Blackmore) who agrees that our minds construct ‘dubious stories whose purpose is to defend a superfluous and inaccurate sense of self.’<sup>57</sup> This skeptical position in regards to both the self and the veracity of the self-story is widespread. Justin Leiber reflects this general skepticism when he writes,

... consciousness tries to perceive and explain what is going on with the much larger unconscious part of our cognitive activity. It is concocting a story, often a very good story, about what is going on, rather than running everything...in a sense my consciousness is just as much or even more a narrative fiction than any

---

<sup>54</sup> Dennett, 1991, p. 407.

<sup>55</sup> Dennett, 1991, p. 428.

<sup>56</sup> Dennett, 1991, p. 428.

<sup>57</sup> Blackmore, 1998, p. 230.

story.... We have probably learned how to have an inner life by listening to stories.<sup>58</sup>

Dennett has been criticized for going too far. You can't just deny that humans are conscious or self-conscious, at least not all of the time. It is that self-conscious experiential quality that is absent from Parfit's views on personal identity, and which make his account so removed from personal experience. I raise these matters in the next sections.

### **3.6 Critique of Dennett, Schechtmann and Velleman.**

I have some general criticisms I want to put forward against the idea that 1) there is a narrative self and 2) that the self is just a narrative. Such criticisms will apply to some extent to all three writers. First, however, I want to address each account directly on its own merits. I want to demonstrate that neither Dennett, Velleman or Schechtmann have fully accounted for our self-phenomenology, nor do their narrative accounts of the self deal with some defining features of selfhood as they appear in that phenomenology. Thus, they have not really come up with a viable alternative to the Cartesian self. In a broader context, this means that the narrative is not adequate as an alternative and that it cannot do all the self-work it is supposed to do. It needs some other self or self mechanisms to underpin it. In Chapters 4 and 7, I explore what these might be.

First, most narrativists do not *argue* that the self is just its narrative. Instead, they appear to assume it without offering an explanation. This means they end up describing the role the narrative plays in a well-developed psyche and the subsequent problems that occur should the narrative go wrong. How or why such a narrative comes into being, why it takes the form it does, or why some aspects of each self-narrative are similar, is left unexplained. Schechtmann and Velleman fall under this category. Neither puts forward a satisfactory account of how the narrative comes into being and who or what is responsible for its construction. Without a viable explanation, the narrative as an explanatory tool loses much of its strength.

---

<sup>58</sup> Justin Leiber, 'Re(ad) me; Re(ad) myself' in *Intertextualities: the 13th Annual Conference on Literature and Film*, Florida State University, 22-24 January, 1988.



Schechtmann claims that the self is a concept embedded in the narrative, which it may well be. However, there needs to be some account of why we have or need such a concept (assuming there is no self driving the process), and how or what creates the narrative in which it is embedded. I would have thought that, as *conscious entities*, we would be aware of ourselves as *persons* or at a minimum as human beings, regardless of whether or not we had a self-concept. Why do we need another concept 'self,' to complicate the narrative? Schechtmann claims she needs a self-concept to get the person-concept off the ground. This means that the person cannot be responsible for creating the narrative because the narrative precedes persons. As she says - no self, no person. Yet she contradicts this position by claiming that the *person* creates her own identity and edits the narrative.<sup>59</sup> It is not clear how this could happen.

If persons can't create their own narrative because they don't exist at this point, it would seem we need to start with something like *self-consciousness* to bind one's identity together. In fact, Schechtmann introduces the engaged subject or subjectivity into the picture, in addition to the narrative. This is supposedly what provides the sense of continuity to the person, the same consciousness providing the temporal link between a past and present person. Although not explicitly stated, it would appear that it is this subjectivity that provides the self-concept which is at the heart of the narrative. It seems that Schechtmann still needs something very like a self to generate that sense of continuous being and to generate the self-concept. Thus, the narrative could be said to capture that phenomenology but it does not create it. In her later works, Schechtmann acknowledges the limitations of the narrative in explicating how it comes into play.<sup>60</sup> She also places a greater emphasis on the pivotal role of *self-consciousness* in constituting personal identity and continuity and in producing our phenomenology. These are aspects of selfhood that are not produced by the narrative.

Like Dennett, Velleman claims that the brain creates the person and an inner self via the narrative, although he does not explain how. He does talk about the *faculty* of causal reasoning and the *faculty* of causal understanding as fill-ins for roles the self normally

---

<sup>59</sup> Schechtmann, 1996, p. 93.

<sup>60</sup> Schechtmann, 2011, pp. 397-416.

takes, probably to remove any hint of an inner locus. One could argue that his ‘guises’ represent specific mechanisms in the brain that ‘take care of’ these self-functions. This would give the various ‘selves’ a neurological realization. However, this would give them more reality than Velleman seems to want. Like most writers who posit a narrative, he proffers little explanation of how it happens.

More importantly, Velleman’s account suffers from too many selves rather than too few. For example, it is not clear why he has to posit multiple temporal selves, given he is not defending the existence of a self as some kind of ontologically distinct temporally-continuous entity. If the self is just a term of reference for various capacities the brain has, as he claims, he doesn’t need an explanation of diachronic identity. Even if he wanted to explain supposed continuity, other writers have claimed that the narrative does this. Its function is to tie the historical person-epochs together into the one story, providing all the continuity the self/person needs.

Velleman, however, has a four-dimensional view of time and, hence, a 4D view of temporal identity.<sup>61</sup> The narrative is episodic, as is the self it represents. This, however, is an ontological position and one I cannot address here. If Velleman is supporting perdurance (objects spread out in time), this is not an argument for denying genuine continuity to the self. It applies to all objects that exist in a commonsense ontology. If the 4D view of time is correct, all objects have multiple temporal parts.

It is also not clear why Velleman persists in calling the various guises ‘selves,’ given that he thinks the word tries to do too much in its original Cartesian form. As mentioned in the discussion of his position, the narrative appears to encompass the other guises, which seems to negate the necessity of presenting them as distinct selves. More confusing still, one of those guises (the self-sameness guise) ends up being subject-like in its own right, a kind of bare or minimal selfhood.<sup>62</sup> Exactly where it fits in the Velleman schema, however, is not clear. The sense of subjectivity or self-sameness guise plays no obvious role in the construction of the narrative and it is not a product of

---

<sup>61</sup> This account is most fully explained in ‘So it goes,’ *The Amherst Lecture in Philosophy*, 2006. <http://www.amherstlecture.org/>.

<sup>62</sup> Velleman, *Self to Self*, 2006, p. 9.

the narrative either. This is an indication of how difficult it is to capture everything to do with our self and its phenomenology within a self-constituting self-narrative.

Velleman's account of the self starts off dismissive of the existence of a single self but ends up positing various kinds of selves, none of which seem to capture what it is like to be the self of the narrative. This confusion reflects the difficulties inherent in removing the self as any kind of entity but still trying to retain its phenomenology. The narrative alone doesn't seem able to do the work. Velleman himself positions the narrative to represent just one guise of the self. His account would make more sense if the self-sameness guise was the basis of the narrative that triggered its evolution. However, I think that move would give the self more reality than Velleman wants to credit it with. At least Dennett does offer an account of how the narrative gets constructed without positing an agent, using the idea of a socially constructed software program.<sup>63</sup>

It is telling, though, that Dennett can only remove agency by dismantling phenomenal consciousness altogether. He has to show that what we think is the case is not the case, that much of what we experience is illusory, including the idea that we experience anything at all. Dennett dismisses the phenomena of the self by adverting to Hume and Parfit. He claims that there is no self to be found other than a collection of experiences and, like them, claims there is no determinate identity. Selves change because their content changes. There is no same-self over time. As a consequence, Dennett does not account for what it is like to be a self-conscious entity, because he claims that there is nothing it is like to be such an entity. For Dennett's account to work, he has to argue that there is no such thing as phenomenal consciousness. He does this by arguing that there is no such thing as consciousness per se. All personal experience is an illusion. This means that, for Dennett, the problematic positing of subjectivity, sense of self or conscious engagement underneath both Velleman's and Schechtmann's narratives can be avoided. He can argue that it is all an illusion.

---

<sup>63</sup> I have debated the viability of this model in detail elsewhere. See 'Dennett's meme machine' in *Dennett and the self*, 2002.

However, this is a very high price to pay to prove that there is no self. It is through our experience of ourselves as *self-conscious* beings that there is a problem of the self in the first place. We also distinguish between and contrast ourselves to automatons and simple reflexive organisms on the basis of that sentience. We recognize that there is a qualitative difference between our lives and the lives of toasters, and our experiences and the experiences of rocks. Dennett argues that the difference is quantitative, rather than qualitative, and that one just needs more complexity organized in the right kind of way. I address his explication in more detail in Chapter 5. Even if what he says were true, it does not show that we are not conscious or that human phenomenology is just an illusion created by some idea or concept of itself, relevant to a particular culture. As Block argued, consciousness is unlikely to be an idea first thought up by the Ancient Greeks.<sup>64</sup> Aspects of our phenomenology are experienced uniformly across diverse cultural and socio-linguistic groups. That commonality becomes apparent in our interactions with other humans and our shared descriptions of phenomenal experiences. Human phenomenology is unlikely to be purely the product of an idea.

### **3.7 General problems with the self-as-narrative account of the self**

I have been pushing the idea that there is more to the self than its narrative and that we can't account for all the features of what we take as evidence of selfhood by just positing a self-narrative. I want to now explore this idea of the narrative in more detail to see if it is at least a viable concept or a useful analogy.

It cannot be denied that we do sometimes tell stories about ourselves and we take those stories to be reasonably indicative of what we have done, why we did it and who we think we are at the time. Likewise, we ask for and get stories from other people that enable us to understand their motivations and actions. This is part of our social interaction with others. We are linguistic creatures and interact via language. Most of the time, we accept the stories we are told unless we have evidence to the contrary, i.e. that the person is lying, exaggerating, has misunderstood something or lacks some relevant

---

<sup>64</sup> This point was raised by Ned Block in 'Begging the question against phenomenal consciousness' *Behavioral and Brain Sciences*, 1992, Vol 15, no 2, pp. 205-206.

knowledge. Strawson calls the view that we engage in story-telling the '*psychological* Narrativity thesis.'<sup>65</sup> This thesis takes it as a given that this is what humans do; they tell stories about themselves.

However, it is not straightforwardly evident that, just because we can tell a story about ourselves, we do always tell stories. Not all events in our lives make it into our stories or are ever expressed, even if they could be. The scratching of my head does not become a part of me only by becoming part of my self-narrative. Nor is it narrative in itself, just because the event has a beginning and an end. This would make narrativity claims trivial and, therefore, unimportant, a point I return to below.<sup>66</sup> So, while we *can* make our lives into a story, it is by no means the case that we *do* as a matter of course, or that we want to, or even that we need to. As Strawson argues, placing too much emphasis on the normative role of narrative in human lives can cause psychological distress rather than benefit.<sup>67</sup>

Despite this evidence to the contrary, some narrativists claim not only that we *do* tell a life self-narrative but also that we *should*. The telling of a self-narrative is both normative and ethically preferable, as both Young and Saver and Ricouer indicate. Not only does the narrative make lives understandable and provide meaning, but the lack of a narrative is a sign of pathology. Schechtman claims that the absence of a self-story means you cannot be a fully-fledged person.<sup>68</sup>

Personhood is dependent on you having a reasonably coherent self-narrative. Strawson refers to this as the '*ethical* Narrativity thesis.'<sup>69</sup> Not only is it preferable to have a self-narrative, but it is essential for being considered a person and accessing the kind of considerations persons have a right to in society. Additionally, a coherent self-narrative is a sign of mental health. Thus, there are strong ethical implications of the narrativist

---

<sup>65</sup> Galen Strawson, *Ratio*, XVII, 2004, pp. 428-452.

<sup>66</sup> Strawson, 2004, p. 439.

<sup>67</sup> Strawson, 2004, p.429.

<sup>68</sup> Marya Schechtman, *The Constitution of Selves*, 1996, p. 93.

<sup>69</sup> Strawson, p.428, 2004.

view of personhood.<sup>70</sup> Personhood depends on the having of a self-narrative, and the quality of the narrative determines how one's mental health is viewed by society. Yet the having of a narrative depends on one's linguistic capacity and one's capacity to construct a cohesive narrative. This is likely to put personhood outside the scope of those with limited linguistic capacities, such as children, mutes, those with an intellectual disability, brain damage, dementia or with just poor communication skills. Strawson claims that it also rules out people like himself who are what he terms 'episodic,' those who lack a strong sense of diachronicity.

Strawson draws a distinction between himself as a person and himself as a mental subjectivity. He claims that one can differentiate between the kinds of persistence conditions that pertain to a person and those that pertain to a 'self' (my scare quotes as I do not want to commit to a Strawsonian self at this point).<sup>71</sup> He argues that, while he may experience his *person* as continuous, this is not necessarily the case with his 'inner' self. He claims that he experiences his self as episodic rather than diachronic because he (himself) has no phenomenological sense of diachronic continuity. He claims that human beings with no phenomenological sense of psychological continuity are unlikely to construct temporal narratives in the way expected within the ethical narrativity thesis.<sup>72</sup> He differentiates between himself as the whole person and himself\* as his mental self.<sup>73</sup> From his own perspective, he claims that he does not view past memories as necessarily his\* because he is not sure that the past events happened to him\* as the same self, even though they may have happened to him as the same person.<sup>74</sup> This means that I\* will not feature in a narrative that talks about past and future events; only I as the person will.

---

<sup>70</sup> I say personhood here because narrativists hold that the self is the narrative and the having of a narrative is what constitutes persons, not selves.

<sup>71</sup> Strawson, p 430. 2004.

<sup>72</sup> Strawson, p 433. 2004.

<sup>73</sup> I discuss this view in more detail in Chapter 6, Strawson's selves.

<sup>74</sup> Strawson distinguishes between two senses of 'I' using the asterisk to represent the inner sense, p 433, 2004.

As an episodic, he claims to focus on the present, to have little interest in viewing his whole life as a single work and does not construct himself as part of an ongoing autobiography. This would place him outside of the ethical considerations espoused by Schechtmann. He would not be a person in anything other than a trivial sense. He would have no self-concept.

Interestingly, Strawson's position is consistent with Velleman's concept of the self. He, too, has multiple momentary selves over time. As a consequence, he claims that the narrative can be episodic and does not need to cohere over a life-time.<sup>75</sup> However, as Strawson argues, this makes the narrativity thesis trivial.<sup>76</sup> The recounting of singular episodes or events does not count as a *self-constituting* narrative. Not just any story will do. It has to have the structure of an autobiography. It has to create a reasonably stable diachronic identity. Taking the above into account, it appears that there are problems with positing the construction of a narrative as a requisite for normality or even, personhood. The narrative, by definition, is a highly specialized linguistic genre. It would seem that this condition is too strong and does not accord with our lives as we know them, except in exceptional circumstances.

This does not mean it is not useful to construct a self-defining narrative nor that constructing a different narrative can be beneficial. However, that does not mean it is what we always do, nor that this is what we should do. It does not mean that this is what we, our selves, are or should be.

### **3.8 Narrative, self-identity and authenticity**

All narrative accounts of the self claim, to a greater or lesser extent, that the narrative captures all the important features of our self-phenomenology. They claim that the narrative fully captures all there is to selfhood, namely its identity and capacities as expressed through the narrative. Let's assume that we all tell a self-story and that this produces a reasonably coherent self-narrative. Unlike Dennett, Schechtmann and Velleman claim that the content of the narrative must reflect real events and consistent

---

<sup>75</sup> This is one of Velleman's deviations from Dennett's account of the narrative self.

<sup>76</sup> See above; Strawson, p 439, 2004.

beliefs and values in order to be seen as coherent by others. Persons are ultimately public entities and engage in numerous social contracts and relationships. According to Schechtmann, such relationships are dependent on the overall coherence of a self-narrative, given what others know of the person. This places constraints on the kind of story one can tell. Schechtmann claims that the nature of the self in the narrative is influenced by socio-cultural expectations of what it is to be a 'good' person or to have a well-defined character.<sup>77</sup> She claims that success in our society depends to a great extent on the robustness or stability of the self-narrative, assuming one exists. Thus, the development of our own story will be influenced by the kinds of things we think are acceptable to express. Schechtmann claims that even if I don't reveal some internal states, they are potentially expressible, given the right circumstances. This means no self-story will be completely accurate or cohere exactly.

This position begs the question of what we mean by 'accurate.' According to Dennett, all 'we' are doing is spinning a tale which may or may not be in line with observable events, and of which there is no fact of the matter about the supposed internal contents of the 'self' expressed in that narrative (i.e. the beliefs one holds, the reasons posited for action). While Velleman and Schechtmann demand a slightly less fictitious story, there is still no inner self here awaiting expression who could vouch for its veracity. So it is hard to see how the idea of accurate or inaccurate would get a foot-hold at all. On the other hand, *potentially* expressible thoughts seem to indicate that there are parts of oneself, one's mental content, that are deliberately kept hidden.

This seems to imply that some thing, the person or 'self,' has some power to discriminate, to make choices about what aspects get revealed or which parts are vetoed to never appear in the narrative. This implies both an internal 'decider' and a distinction between its public persona and the content of its inner self. The inner 'decider' would need to be quite sophisticated as they would be making moral and social judgements about what the 'public' can deal with and about what oneself could handle in relation to the public's knowledge of oneself. These are tasks we generally take for granted and ones we would normally allocate to our 'self.'

---

<sup>77</sup> Schechtman, 2005, pp. 9-22.



The issue of genuine self-identity in relation to the narrative comes to the fore in Davies and Harre's paper on the discursive self.<sup>78</sup> I do not have the space to address all the issues in detail, but I want to demonstrate a natural consequence of a narrative self that has no determinate identity and no determining qualities. According to Davies and Harre, '[o]ne lives one's life in terms of one's ongoingly produced self, whoever might be responsible for its production.'<sup>79</sup> For Davies and Harre, the self is nothing more than the manifestation of certain identities or 'positions' that unfold in different discursive practices. Davies and Harre go on to claim that,

each of these possible selves can be internally contradictory or contradictory with other possible selves located in different story lines... Like the flux of past events, conceptions people have about themselves are disjointed until and unless they are located in a story.<sup>80</sup>

What they are doing is setting up a model of personal or self-identity that assumes that all identities and every aspect of our identity are infinitely variable and open to choice. There is no re-identifiable self, there is nothing continuous about oneself; there is just an infinite possibility of selves. The self-narrative takes on whatever identity one temporarily decides to adopt at that time in that situation, a problem I identified with both Schechtmann's and Velleman's accounts.

The problem with this account of the self-narrative is that it is ultimately unviable as an account of the self or even one's personal or self-identity. Given the fluidity and instability of this type of self-construction, and the eminent variability of the self that is supposedly manifested, it seems to matter little how, who, or as what we are positioned. It assumes that all relations are open to change and in need of redefining and that this is possible. Yet any repositioning would be temporary and relevant only to that particular discourse. It is also unclear how one could establish a benchmark for the ideal or most acceptable position to be in. What would be the normative or most ethical/just/fair arrangement and why? No position is any more moral or more meaningful or even more

---

<sup>78</sup> Bronwyn Davies and Rom Harre, 'Positioning: the Discursive Production of Selves' at [www.massey.ac.nz/~Alock/position](http://www.massey.ac.nz/~Alock/position), 1999.

<sup>79</sup> Davies and Harre, 1999, p. 4.

<sup>80</sup> Davies and Harre, 1999, p. 4.

justifiable than any other *because* no person has any genuine identity nor any self beyond one's transient identity.

This discursive, narrative view assumes that there are no deeper facts to identity and/or selfhood other than situational dynamics. Unfortunately, one cannot be guided in one's discourse by one's true nature, or one's desired nature, or even by the need to express oneself in a genuine interchange with another discussant, because there *is nothing it is like* to be you or them. The 'self' in this account is empty of genuine content or filled by (temporary) artifices collected from past discourses. This doesn't even work as an account of personal identity. It is far removed from the 'stories' most people tell about themselves in relation to who they are. The problem is that in order for one to strive for change in oneself, any change, there has to be something residual to change and there has to be something one wants to change to. Ideally, what one strives for would be an improvement on what (character) one has or is, or one strives for a more authentic expression of one's self. However, in this account, there is nothing that one is. There is no essential or other type of self that is in need of expression, so any account of who one is, is as good as any other. There is also no ideal self that one could strive to achieve; there is no ideal positioning.

This account does not reflect our social or phenomenological reality. That phenomenology does not, in itself, assume an essential self or a fixed personal identity, but there do appear to be genuine constraints on what we can do and who we can be. And we don't just experience ourselves as the way we appear in any one discourse, nor do we experience ourselves *because* we appear in a discourse. Ultimately, what this version misses altogether is the person or self that is engaged in the discourse. It misses the entity that is capable of repositioning itself and who is responsible for the reconstructing of that self-story.

Like Davies and Harre, many post-modern literary accounts of the self assume that the self is 1) a social construction, 2) fully explicable by its identity and 3) capable of infinite reconstruction. There are no deeper facts of the matter to the self's existence.

The self's subsumption under the rubric of identity is taken as a given.<sup>81</sup> This position is also reflected within many psychological accounts of the self as well.<sup>82</sup> As Polkinghorne states, '[o]ne function of psychotherapy is to assist in the reconstruction of a meaning-giving narrative of self-identity.'<sup>83</sup> At the same time, however, the same writers are concerned about the 'authenticity' of the narrative and the role of 'false' selves in mental health. They claim that selves need to be authentic and represent the 'real you.' Too much deviation between a false self and one's real self can lead to complexes, psychoses and neuroses, the result of living an inauthentic life as a false self. Winnicott believed that the 'false self has as its main concern a search for conditions which will make it possible for the True Self to come into its own.'<sup>84</sup> So while the construction of one's identity or self is an open-ended exercise, there can be conflict between this construction and what your real nature is or would have you be.<sup>85</sup>

I am not going to spend any more time on this aspect of narrative theory as it just seems confused. However, this kind of contradiction is often left unexplored. Needless to say, it makes no sense to talk of either a 'real' self or an 'inauthentic' self within the framework of the narrative. If there is an authentic expression of oneself that is possible, which can generally be discovered during therapy, then there is some genuine nature or identity to that self.<sup>86</sup> Yet the whole basis of the narrative account rests on there being no residual or genuine identity. It is all construction. There is no inner self or identity that needs expression, such that its denial will lead to psychosis. Yet the preponderance

---

<sup>81</sup> Accounts are too numerous to mention but see J. Bruner, *Making Stories: Law, Literature, Life*, Cambridge, MA: Harvard University Press, 2002.

<sup>82</sup> See i.e. J. Bruner, 'The Narrative Construction of Reality,' *Critical Inquiry*, Vol. 18, No. 1, 1991, pp. 1-21; Donald E. Polkinghorne, 'Narrative and Self-concept', *Journal Of Narrative And Life History*, 1(2 & 3), 1991, pp.135-153.

<sup>83</sup> Polkinghorne, 1991, p. 135.

<sup>84</sup> Daniel Winnicott, *True and False Self*, 1960, p. 96. On the other side, John Simons claims that Foucault took issue with the concept of a "true self" on the grounds that the self was a construct, not (as in the Romantic paradigm) an essential to be uncovered: he maintained that because "the self is not given to us...there is only one practical consequence: we have to create ourselves as a work of art". John Simons, 'Michel Foucault,' in John Simons (ed.) *Contemporary Critical Theorists: from Kant to Lacan*, Edinburgh University Press, 2006, p. 196.

<sup>85</sup> Most often manifested as repressed sexual orientation.

<sup>86</sup> Discussions like this can emerge in talk of MPDs/DIDs. One particular self or identity is considered to be the 'real' one. I discuss this in the next chapter.

of identity issues and supposed conflict between societal-parental expectations and one's 'true' nature, frequently discussed in the psycho-analytic and fictional literature, lead one to the opposite conclusion.<sup>87</sup> According to the tenets of narrative psychotherapy, living an inauthentic life that belies one's 'true' self will have a negative impact on one's mental health. This evidence runs counter to the claim that the self is just a fictional narrative that can be re-constructed how we wish, whoever the 'we' is.

### **3.9 Zahavi, self-phenomenology, and the role of ipseity in selfhood: the limitations of the Narrative self**

Despite the issues raised above, it seems a given that we are not born with a fully-fledged identity. As babies and children we begin to develop into the kind of person we will become. Exposure to influences and events in our immediate environment both shape our development and trigger our dispositions. It is still debatable how many of our beliefs, personality traits and talents are the result of our genetic make-up; we are not born an empty slate. Nevertheless, we have an emergent or developing identity that cannot be predicted and which will be driven as much by environmental factors as by our genes.<sup>88</sup> In this sense, we do learn about ourselves (at least our capacities) as others do, by how we react and adapt to new and challenging circumstances. We can be pleasantly surprised and disappointed. This new information may require us to adjust how we think about ourselves, to *revise* our stories.

The story we tell, if we do, is about our personal identity; who we are as a person. Narrativists want to go much further. They claim the narrative is all there is to the self and that this fully explains all our self-phenomenology. They argue that there is no other self than the one identified in the narrative and that this self only exists by virtue of the narrative. Thus, the narrative creates the belief in us that we have a self (a self-concept), that that self is singular or unified and that that self is capable of agential actions. In fact,

---

<sup>87</sup> See Alison Bechdel, *Are you my mother?* HMH, 2012, USA, pp 77-110 as an illustration of the self as constructed by the other, yet unable to express its real nature.

<sup>88</sup> In *Rethinking Innateness: A connectionist perspective on development*, the authors claim 'Genetic constraints interact with internal and external environmental influences to produce a phenotypic effect.' (preface, p. xii), Jeffrey L. Elman, Elizabeth Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi and Kim Plunkett; Massachusetts, MIT Press, 1998.

narrativists are claiming two separate things here – 1) that self and other knowledge is always mediated through the self-narrative and 2) there is nothing more to that self than the existence of that narrative. In this section, I want to dispute both claims.

In Chapter 2, I argued that personal identity proponents like Parfit ignored our self-phenomenology. For them, there is nothing it is like to be a self. Experiences are subjectless. Dennett argues that phenomenal consciousness itself is an illusion created by an idea, or a meme-plex of ideas.<sup>89</sup> While Velleman and Schechtmann do not deny subjectivity, they do deny a subject; subjectivity is just the capacity to access one's memories. However, I have argued that both authors rely on that subjectivity to provide a core consciousness or self-consciousness without which the narrative would not work. Shoemaker argues that the use of 'I,' the traditional denoter of a subject, is immune to error when referring to self-experiences, such as pain.<sup>90</sup> While the immunity to error principle has been challenged, given one could be mistaken about the nature of the experience (i.e. think one is in pain when one is not), one is never mistaken about who is having the experience or that there is an experience.<sup>91</sup> This would not be possible if there were no subject to experience it. It seems highly unlikely that experiencing a pain relies on its place in a narrative before one is aware of it, even though the narrative may change how the experience is viewed retrospectively.

Zahavi claims that Sartre, famously taken to argue against a self, posited a 'fundamental self-givenness or self-referentiality' which he called 'ipseity.'<sup>92</sup> This state, which I have referred to in Chapter 2, has been posited by many later authors as the basic sense of 'being someone or some thing.'<sup>93</sup> Merleau-Ponty famously refers to it as 'being-in-the-world' to capture the sense of embodiment that runs counter to the Cartesian mental

---

<sup>89</sup> I address Dennett's position in more detail in Chapter 5.

<sup>90</sup> Sydney Shoemaker, 'Self-reference and self-awareness', *Journal of Philosophy*, vol. 65 no.19, 1968, pp.555-567.

<sup>91</sup> Uriah Kriegel discusses the challenges to this view under 'Epistemic peculiarities of self-consciousness' in the Internet Encyclopaedia of Philosophy, 2007. <http://www.iep.utm.edu/self-con/>

<sup>92</sup> Dan Zahavi refers to 'Being and nothingness' in 'Self and other: the limits of narrative understanding,' D.D. Hutto (ed), *Narrative and understanding persons*, Cambridge, Cambridge University Press, 2007, p.183 of pp. 179-201.

<sup>93</sup> I have already mentioned Ricouer, Gallagher and Damasio. Zahavi also cites Metzinger but I would disagree with his inclusion as he denies phenomenal consciousness along similar lines to Dennett.

self.<sup>94</sup> This sense of self is more than just having an experience of something, like the taste of curry or the smell of fresh grass. Zahavi characterizes it as being ‘immersed in conscious life.’ He goes on to say that,

the (minimal or core) self is claimed to possess experiential reality, it is taken to be closely linked to the first-person perspective, and is in fact identified with the first-personal *givenness* of the experiential phenomena.<sup>95</sup>

Importantly, while individual experiences of particular things can vary across people, the sense of self remains constant. It is the sense of ‘mineness,’ that these things are being experienced by me, that ‘I’ am the consciousness that is engaging with them. This ipseity is fundamental to the sense of self yet it is ignored by the narrativist, hence also by the narrative. While the ‘I’ of the narrative could *claim* to be phenomenally conscious like any robot could (this is Dennett’s position), this would not explain the uniform and consistent features used to describe ipseity. It would also reduce all conscious experience to merely ‘words wanting to be said.’<sup>96</sup> As I have stated previously, advocates like Dennett need to show that we are *not* phenomenally conscious rather than the onus being on me (for example) to show that we are. Ultimately, it is this sense of oneself as a something that is alive and experiencing this event, that grounds the sense of self and, hence, the belief in a self. It is a phenomenal experience of being distinct from others, of being singular and of having boundaries. These properties represent a core self which would also be present in non-linguistic and pre-linguistic humans. While the narrative may change, these properties remain constant. As Zahavi says, the self should be considered a ‘pre-linguistic pre-supposition for any narrative practice.’<sup>97</sup>

Zahavi raises two other related problems with the narrative account of the self.<sup>98</sup> He says that narrativists assume two things, both of which he disagrees with. These are that:

- The narrative is the primary access to the self, the source of all self-knowledge

---

<sup>94</sup> Maurice Merleau-Ponty, *Phenomenology of Perception*, Routledge and Kegan Paul, London, 1962.

<sup>95</sup> Zahavi, 2007, p. 183.

<sup>96</sup> Dennett, 1991

<sup>97</sup> Zahavi, 2007, p. 184.

<sup>98</sup> Zahavi, 2007, pp. 179-201.

- Our relations with others are mediated through their narratives.

The first claim relates to the epistemological claim that I am arguing against, that the self can be fully captured (and understood) by its narrative. Zahavi calls this the ‘exclusivity claim.’ In a similar vein to Shoemaker, Zahavi claims that the existence of a self-narrative actually presupposes a pre-lingual ‘core’ self. One cannot claim experiences as one’s own without first having a sense of oneself. To do so requires a first-person perspective. This grounds the later use of ‘I.’ There is a growing body of experimental evidence to support this view. Research findings indicate that very young, *pre-lingual* children can differentiate self from other and that they appear to have a rudimentary sense of self.<sup>99</sup> Such a sense of self is obviously not a product of a social narrative.<sup>100</sup>

Zahavi claims that this indicates that we gain self-knowledge through other means besides the narrative. A first-person perspective emerges from the self-identifying effects of (conscious) perception. We do not gain access to our memories, our experiences or our mental capacities merely through the narrative. These find expression in the narrative, not the other way around. This is why there may be parts of our self that are not always verbally expressed. Being expressible does not make them part of a narrative. Either we learn about ourselves through the narrative because that is all there is to our self or we do not (at least not all of the time). The fact that there are other ways of learning about ourselves indicates that the narrative is not the sole source of self knowledge, indicating there is more to the self than is expressed (or expressible) in the narrative.

The second related problem Zahavi raises is an interesting one. According to narrativists, we learn about others and understand them through their narratives, just as

---

<sup>99</sup> See ie Meltzoff and Moore, ‘Infants’ understanding of people and things,’ in José Luis Bermúdez, Anthony J. Marcel and Naomi Eilan, (eds), *The body and the self*, Cambridge, MIT Press, 1995, pp. 43-69; E. Fivaz-Depeursinge, N. Favez and F. Frascarolo, ‘Threesome intersubjectivity in infancy.’ *The structure and development of self-consciousness: Interdisciplinary perspectives*, D. Zahavi, T. Grünbaum and J. Parnas (eds.), Amsterdam, John Benjamins, 2004, pp. 221-34.

<sup>100</sup> Ian Ravenscroft says that a pre-lingual narrative could exist using mentalese, Fodor’s LOT (private discussions). Even if true, this would still not support the claim that the self is just its narrative, nor that it is socially acquired, and language per se is not always narrative in form. If mentalese is innate it might argue against the claim that the self is a story.

they understand me through mine. Dennett terms this the ‘heterophenomenological’ or third-person approach, which I have referred to before. However, this is not quite accurate. First, as Zahavi points out, intersubjectivity begins prior to the development of a narrative. Again, very young children show a remarkable capacity for interacting with others that indicates awareness of the other as an agent like themselves, not as another (inanimate) object. They tend to ignore dolls, teddies and even live animals in their environment but respond immediately to the presence of other humans.<sup>101</sup> This tendency is pre-lingual and becomes sophisticated very quickly, long before the capacity to construct a meaningful narrative emerges. One response to this has been to claim that these interactions qualify as quasi-narratives because actions have a structure similar to the narrative form (beginning middle end). This rather ad hoc justification means that anything can qualify as a narrative, even simple experiences or actions, as they are necessarily temporal.

However, this is what Strawson terms the trivial version of narrativity. Just because someone could tell a story doesn’t make it a self-defining story. It is not self-constituting. If the term ‘narrative’ becomes too broad it ceases to have any real meaning and this weakens its capacity to be explanatorily useful. It becomes empty or a cliché. Research indicates that infants do have a real sense of self/other differentiation that needs some kind of explanation. We can’t just dismiss it as either pseudo-intersubjectivity or claim it is an example of narrative, but in a pre-narrative/ pre-lingual form. For the narrative to be self-constituting, it has to be robust.

Secondly, it is not obvious that we understand and relate to others just by drawing inferences from their narrative. The age-old problem of positing the existence of other minds exists because we do not have *direct* access to another’s experiences and hence, have no direct evidence that they have minds like our own. Consequently, we assume our understanding of the other is always mediated by their behaviour, their stories and the inferences we draw from these. However, emerging developmental evidence from very young children seems to suggest that this is not the whole story. There is a sense in which young children recognise another human in their field of vision as an *agent* of

---

<sup>101</sup> Meltzoff and Moore, 1995.



their actions and as a *possessor* of internal states or experiences.<sup>102</sup> It is a kind of perception like all perceptions. Often it is this capacity that is taken as lacking in people with autism.<sup>103</sup> Very young infants have no narrative and lack access to our narratives. Yet they respond to other humans *as if* they are goal-directed or can cause things to happen.

Likewise, we recognise in the other a sentient being who we cannot know in their entirety. We can and do impart meaning to the gestures and actions of others because they are the actions of *other sentient intelligent beings*, not just because they are figured in a storyline. This is despite their sometimes limited linguistic capacity, story-telling skills or reluctance to communicate. Similarly, we recognise in others that there is something about them that we cannot access or know; we don't know what it is like to be them from the inside. We recognise a subjectivity like our own with all that that entails. This is not given in the narrative.

What all of this indicates is that the narrative does not do the work it is intended to or believed to do. It does not capture or explain the conscious sense of self or being that is an integral part of our belief in a self and which is likely to ground that sense of self. We experience ourselves from a position of selfness. The only way to deny that is to deny phenomenal consciousness altogether. As a consequence, not all our self properties are the result of the having of a self-narrative. Much of what we experience as part of our sense of self is left unaccounted for, both here and when self is reduced to just identity.

### **3.10 Conclusion – the self is not its narrative**

It is clear that the self-as-narrative thesis is flawed. It does not capture all the features of our selfhood nor our interactions with others. While the narrative concept might help us to understand an important feature of language use or even the power of language to influence human behaviour, it is inadequate as an explanation for the existence and

---

<sup>102</sup> Meltzoff and Moore, 1995.

<sup>103</sup> L. Q. Uddin, M S Davies, A A Scott, E Zaidel, S Y Bookheimer, m Iacoboni, M Dapretto, 'Neural Basis of Self and Other Representation in Autism: An fMRI Study of Self-Face Recognition,' *PLoS One* vol 3, issue 10, e3526. doi:10.1371/journal.pone.0003526, 2008.

nature of the self. Narrativists assume that the self is socially constructed.<sup>104</sup> It can only be acquired in a social context, in conjunction with others. Narrative advocates claim that the very concepts I use to describe myself or my capacities will be dependent on the society in which I live and the historical period I occupy. They depend on tradition and current theory, and the story-telling conventions of our day.<sup>105</sup> Our ‘selves’ enter the public domain where they are constructed to match the expectations of the culture or society we inhabit. ‘I’ am then the product of the narrative with all its real and fictive qualities.

While our personal identities may be socially constructed to a greater or lesser extent, there is not the same evidence to support claims that our conscious capacities or sense of self are also socially constructed. Claims of selfhood do not seem culturally dependent. According to Strawson, there is little difference in the way humans claim to experience themselves as self-conscious entities across cultures.<sup>106</sup> The dependence on language also limits the existence of the narrative, and hence the self and possibly personhood, to those with a linguistic capacity. This has limitations on those who can and can’t be said to have a self/be a person. It is particularly problematic with pre-lingual humans such as children, those with severe disabilities and some deaf mutes. One could deny them personhood, but that does not entail one can also deny them selfhood.

At the same time, the demands of the autobiographical linguistic genre are too strong. The narrative can only construct the self as it appears to exist in our phenomenology if it follows strict guidelines. Not any story will do and our sense of self is quite robust. The narrative has to literally be autobiographical in every sense of the word. As we have seen, this is not the way we live and it excludes too many humans who could not or do not comply.

From the analysis presented here, it seems that the narrative requires a pre-existing self concept of some form to generate a self narrative. It needs a first-person perspective. Yet

---

<sup>104</sup> Dennett’s meme machine is the ultimate social mechanism.

<sup>105</sup> Polkinghorne, 1991, pp.135-6.

<sup>106</sup> Strawson, 1998, p. 2.

the self-as-narrative thesis ignores phenomenal consciousness and our basic sense of self. That basic experiential sense of existing provides many of the properties that make us feel like we are selves. Evidence from research into early childhood poses difficulties for the theory. Because the narrative does not capture all facets of the self it is not the only means of acquiring self-knowledge. We find out about some aspects of ourselves through direct experience, unmediated by the narrative. Self/other relations are also not solely mediated through the narrative. We sense other people as active subjects. We also do not assume that their narratives capture exactly what it is like to be them from their perspective. This is something we know we can't know exactly, regardless of how detailed the narrative is. We accept others as *subjects* of experience, not just as objects in a narrative. Most importantly, the narrative model seems to work best by ignoring our phenomenology and denying self-consciousness. This seems a radical step for very little gain.

I have argued that the narrative account of the self, the view that the self is just its narrative without remainder, is unviable. It fails to do the work it is set up to do. It does not account for the self or our self phenomenology. I do not think that the narrative is a fundamental part of our cognitive make-up. We do have an identity, both a personal identity and a numerical identity, but I am not convinced we also have a narrative identity. Nevertheless, we do have the capacity to tell a self story, should we need to.

Similar to Zahavi and Damasio, I have intimated that there is a basic or core self that has the capacity to ground our identity and which is essential for personhood.<sup>107</sup> This core self will be explored over the next few chapters.

In the next chapter, I explore a range of neurological pathologies that have been identified as 'neuropathologies of the self.' These pathologies have been referred to in detail by numerous authors to demonstrate that there is no such entity as the self. They have been taken as evidence to show that there is no inherently singular or unified self

---

<sup>107</sup> See A. R. Damasio, *Descartes' Error: Emotion, Reason and the Human Brain*, London, Papermac, 1996 and *Self Comes to Mind*, Pantheon, New York, 2010. I discuss his view in detail in Chapter 7. Damasio demonstrates through neuropathology that the narrative self is dependent on the core self. Damage to the fundamental mechanisms that produce the sense of self will cause disintegration of the self. Damage to higher order mechanisms that produce the narrative self do not.

that could play the role of the Cartesian self. I address this evidence and argue that it does not demonstrate that there is no self, other than the non-physical Cartesian self. In fact, I argue that it demonstrates that there are fundamental mechanisms in the brain that have specific self-functions. When these are damaged, the sense of self is likewise damaged. This adds additional support to the case against the self-as-narrative. It also suggests a provisional model for a self-like entity or self-system.

## CHAPTER 4

# THE SELF AND ITS PATHOLOGIES

Pathology has made us acquainted with a great number of states in which the boundary lines between the ego and the external world become uncertain or in which they are actually drawn incorrectly. There are cases in which parts of a person's own body, even portions of his mental life...appear alien to him and as not belonging to his ego; there are other cases in which he ascribes to the external world things that clearly originate in his own ego. Thus even the feeling of our own ego is subject to disturbances and the boundaries of the ego are not constant.<sup>1</sup>

### 4.1 Introduction- the issue of unity

In the last two chapters, I have argued that what we generally take to be the self, the kinds of properties and capacities that fall under the term 'self,' is not fully captured under the rubrics of personal identity or narrative identity. Both positions subsume the self under matters of identity. While the narrative self can be considered distinct from one's person, insofar as the narrative grounds or precedes personhood, the self of the narrative is still a construction *of* the narrative. It just provides the 'self' concept or idea of a self. There is no deeper fact of the matter. I have argued that treating the self as just its social or personal identity does not address the self-phenomenology, the sense of self or the quality of ipseity. Constructing a narrative identity does little to address this, other than by including the phenomenology in the narrative. It leaves too much unexplained (in relation to who/what constructs and edits the narrative) and ignores too much (in relation to how we are/act in the world). This inevitably leads to the conclusion that the self-phenomenology must be illusory. I have argued that narrativists either need to deny the existence of both phenomenal and self-consciousness, or they need to posit a core self to make their story work. Denying phenomenal or self-consciousness removes the problem of the self by avoiding the hard questions. It does not solve the problem of the self.

In this chapter, I want to examine a slightly different approach to the issue of the existence of a self. I want to look closely at a collection of neuropathologies that appear

---

<sup>1</sup> S. Freud, *Civilization and its discontents*. Standard edition, Vol. 21, London, Hogarth Press, 1930, p. 66.

to attack the very self-phenomenology that I argue is at the heart of our sense of self. If the supposed 'self'-phenomenology is not uniformly experienced by all cognising humans, and if it can be shown to be false or delusional, then this casts doubts on the reliability of our claims about that phenomenology. If we can be mistaken about ourselves and the extent of our agency, then this would argue against the existence of a genuine agential self responsible for that phenomenology. This is particularly the case with the self-properties of singularity, unity and continuity over time.

It is the purpose of this chapter to examine a selection of such self-disorders to see what can be inferred in relation to a 'normal' self-phenomenology, the nature of the self and its potential existence. In the process, I hope to demonstrate that some or all of the following are true or, at a minimum, are not incompatible with the phenomenology associated with self-pathologies. This would serve to weaken the evidence of psychopathologies for the non-existence of the self.

- Some pathologies are ailments of personal identity rather than ailments of the self
- Selves are not as fragmented or as incoherent as they can appear
- Apparent fragmentation of the self does not imply an inherent lack of unity
- Self-phenomenology has a biological or neurological basis as well as a socio-cultural one
- Selves can be damaged and even destroyed

I argue that the neuropathologies of the self do not, in themselves or collectively, demonstrate that there is no self, nor that the self-phenomenology is illusory. I claim further, that the kind of damage that results from the pathologies demonstrates the existence of very specific self-mechanisms in the brain that appear to work together to create the self-phenomenology and our self-capacities. This argues against the role of the narrative in constructing all our self-phenomenological experiences. I claim that the

concerted actions of these self-specifying mechanisms could be said to constitute a core self, similar to that posited by Damasio.<sup>2</sup>

I also re-iterate an important point I raised in Chapter 2. The damage or destruction of the self does not belie its existence, just as the damage or destruction of any object does not belie its existence. No more nor less than any existing physical object, selves will be vulnerable to destruction in certain circumstances. The self-pathologies addressed here seem to be such circumstances.

## **4.2 How pathologies have an impact on the self/ the sense of self**

In the normal course of development, each human being develops a sense of who they are as a particular kind of psychosocial entity. This perception generally entails an acceptance of oneself as a singular agent capable of acting cohesively in the world. One has a sense of oneself as a particular kind of person (one's qualitative identity), and a sense of oneself as a certain kind of cognising entity, (i.e. that one is more or less capable of thinking, reasoning, making decisions and acting for oneself). Unless one is born with a neurological disorder or physical disability, one has an expectation that one's brain and body will work together cooperatively to perform one's intended acts. Except in exceptional cases, I assume that I can get my body to perform my will and that I know what my body can and can't do intimately because it is a part of me, even if this capacity is sometimes over and under-estimated. In 'normal' circumstances, my arm reaches for the cup in conjunction with my decision to drink, in a seamless action. I do not question my arm's action because it coincides with my decision to act. My mind and body appear to be as one unit.

However, there is evidence that this is not always the case and that it need not be the case. There are a variety of psycho-social and neuro-pathological conditions that can undermine the expected sense of unity and oneness one normally experiences between one's mind and body, and which can even appear to break that link at times. The loss of

---

<sup>2</sup> Although discussed in his earlier works, this idea is most clearly expressed in Antonio Damasio, *Self Comes to Mind*, Pantheon, New York, 2010.

that link has the capacity to create a sense of alienation from one's own body or even from parts of one's own mind. This dissociation and lack of cohesion can lead to a breakdown in the phenomenal sense of self. Sufferers claim to no longer experience themselves as 'themselves' or as 'singular.' In extreme cases, the person with the pathology may manifest bizarre or uncommon behaviours, attitudes and beliefs about themselves, about their relation with the world and about others. The experience of 'James' who suffers from schizophrenia, illustrates both that loss of agency and the bizarre beliefs that can manifest as a result. The passage below gives an indication of how horrendous schizophrenia must be for those who are afflicted. He describes his experience, *as he lives it*, as follows:

At some point the parasite had bitten into...my inner brain. Now the instructions are to bite off my penis, slash my face, jump in front of a car. If these instructions are not carried out, I will be forced to comply. I am helpless and worthless. The parasite is a source and a voice, in the depth of my brain, in the flesh of my leg and hands, in the curve of my back, with terrible control over what I am allowed to think and do.<sup>3</sup>

Unlike our everyday experience of ourselves, James is not alone in his head. The presence of demanding, alien voices leads to an experience of a loss of control over his actions. This has a negative impact on his self-phenomenology and his experience of himself as a free agent. He goes so far as to claim that death would be preferable to his current existence.

An examination and understanding of the cause of psycho-neuropathological ailments like schizophrenia should give us considerable insight into the normal functioning of our bodies and brains, as well as some understanding of the self and its phenomenology. Our sense of self is mediated by our perceptions. The kind of impact that neurological or physiological illnesses have on our perceptions of ourself and the world we live in should help us understand what and how our phenomenology is created. James' comment above, seems to indicate that the traditional sense of self, along with its relation to its environment, can be severely challenged and even permanently damaged by schizophrenia. This indicates that the sense of self is vulnerable to disintegration if

---

<sup>3</sup> W. Heinrichs, *In search of madness*, OUP, New York, 2001, p. 5.



certain neurological or other events occur. In a similar vein, there is evidence that some life-changing events, such as extreme physical and emotional trauma, appear to cause a splintering of the self or the creation of more than one self. Multiple personality disorder (or DID as it is now known) is a case in point. This would also indicate that one's environment or what one experiences can make or break the self.

These phenomena have led many philosophers to draw sceptical conclusions about the existence of a self and to cast doubt on its phenomenology and our reporting of that phenomenology. They have led to the idea that the self is not the single enduring entity it appears to be and that its nature is a product of its social environs. Dennett uses the existence of dissociative identity disorders (DIDs, formerly MPD) to demonstrate that there is no Cartesian or necessarily intrinsic *singular* self. He claims there is no reason that selves need to be 'All or Nothing and One to a Customer.'<sup>4</sup> Metzinger refers to DIDs, anosognosia and phantom limbs to demonstrate that the phenomenal self and our phenomenological claims about our self, are a delusion.<sup>5</sup> Other writers use schizophrenia as an indicator of multiple agency or lack of agency altogether.<sup>6</sup> It would seem that, rather than informing us about the self, certain neuropathologies seem to argue against its very existence.

In the following sections. I investigate self-pathologies like schizophrenia, DID/MPD and Cotard's Syndrome to examine what is known of the causes, and to assess the behavioural outcomes and symptoms as they relate to the self and the sense of self. I will draw implications from this exploration about the nature of the self and how the sense of self is created. I will demonstrate that the implications of neuropathologies for the nature of the self are not as straightforward as they appear. There is still a great deal that is

---

<sup>4</sup> Daniel Dennett, *Consciousness Explained*, Little, Brown, Boston 199, pp. 420-425.

<sup>5</sup> T. Metzinger, 'Neurophenomenological Case Studies II,' *Being no-one*, MIT Press, 2003.

<sup>6</sup> See E. Daprati, N. Franck, N. Georgieff, J. Proust, E. Pacherie, J. Delany, M. Jeannerod, 'Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients,' *Cognition* 65, 1997, pp 71-86; W. Heinrichs, *In search of madness*, 2001; G. Caruthers, 'A Model of the Synchronic Self,' *Consciousness and Cognition*, Vol 16, 2007, pp 533-550; S. Harter, *The Construction of the Self: A Developmental Perspective*, Guilford Publications, 2001.

unknown about many psycho-social disorders like schizophrenia and DID.<sup>7</sup> Their aetiology is inconclusive, it is not always clear what or why things have gone wrong or why they manifest the bizarre symptoms they do.

I conclude that psycho-neuropathologies like schizophrenia and DID do not argue against a singular or unified self as claimed and that there is strong evidence to show that, in most cases, the self-pathological phenomenology can be traced to neurological damage or a chemical imbalance in the brain. The role of specific neural substrates in producing our self-phenomenology would support a biological basis to the sense of self. This would argue against its supposed illusory status and its dependence on social norms and beliefs. It is likely we have no choice but to experience the phenomenology we do. One could also infer that those self-functioning mechanisms responsible for the phenomenology could warrant the title of ‘self.’

### **4.3 Schizophrenia and delusions of control**

Schizophrenic sufferers are seen as the classic representation of madness.<sup>8</sup> They are traditionally portrayed in literature and film as irrational, unpredictably violent, living in their own ‘bizarre’ world, talking to themselves or to imaginary characters and often as having a ‘split’ personality. The last inference is drawn from one of the most marked symptoms of schizophrenia, what has been termed the ‘delusion of control,’ the supposed presence of an alien voice in their head or the experience of some other agent beside themselves who takes control of their actions or their will. The comments by James, above, are a good illustration of the bizarre nature of the delusions. According to DSM IV, a diagnosis of schizophrenia involves the presence of two or more of the following symptoms from criteria A:

[A] Characteristic symptoms: ... each present for a significant portion of time during a 1-month period (or less if successfully treated):

(1) delusions

---

<sup>7</sup> See Heinrichs, 2001 and D. Papineau, ‘Mental Disorder, Illness and Biological Disfunction’ *Philosophy* 37, 1994, pp 73-82.

<sup>8</sup> Heinrichs, 2001, Chapter 1.

- (2) hallucinations
- (3) disorganized speech (e.g., frequent derailment or incoherence)
- (4) grossly disorganized or catatonic behaviour
- (5) negative symptoms, i.e., affective flattening, alogia (poverty of speech), or avolition (lack of motivation)<sup>9</sup>

While there are other criteria such as social dysfunction, mood disorder and duration that can contribute to a diagnosis, most weight is placed on the presence of what are termed ‘positive’ symptoms or those listed from (1) – (4). Most often, if the first two criteria from A are present in enough strength, especially if the delusions are bizarre and the hallucinations involve hearing voices or engaging in dialogue with imaginary perpetrators, then the patient is considered to be suffering from schizophrenia. In fact, claiming either lack of ownership of thoughts (alien thought insertion or that voices in the head are not their own) and thereby a lack of responsibility for the resulting actions, and extending claims of self responsibility to actions that are not their own, are considered defining conditions for schizophrenia.<sup>10</sup> This is why it is the paradigmatic representation of madness. The sufferer claims to hear voices no-one else hears and seems to interact with people (or entities) no-one else can see. They can be deluded and act as if they are God, the Pope or an avenging Angel.<sup>11</sup>

Schizophrenic patients suffer from what has been referred to as ‘Ich störungen,’ a disturbance of the experiencing ‘I’ or what Freud called ‘ego’ boundaries.<sup>12</sup> This type of delusion or hallucination creates a different phenomenal experience to that of most cognizing drug-free humans and it has an impact on their experience of themselves as a unified agent. From what they say, their internal, mental world is no longer just the province of themselves and their own thinking. Some other presence seems to have access to or seems to have control of what should be a sacrosanct place, their own mind. This means that the property of ‘mineness,’ related to feelings of ownership, identity and

---

<sup>9</sup> DSM IV Manual, <http://www.dnalc.org/view/899-DSM-IV-Criteria-for-Schizophrenia.html>, accessed July, 2012.

<sup>10</sup> DSM IV.

<sup>11</sup> Heinrich, 2001, p. 4, William ... ‘If only they would do EEG and IQ tests I could prove that I am God.’

<sup>12</sup> Metzinger, 2003, p. 438.

control, is not as robust or as inviolate as it once was. The boundaries between self and the rest of the world appear to have been transgressed.

Given the unlikely occurrence of alien possession or talking parasites, schizophrenia has been categorised as a case of inaccurate *self*-ascription.<sup>13</sup> It has also been cited as evidence for the non-intrinsic singularity of the self and a lack of capacity for reliable self-knowledge. These are two of the most intrinsic features of selfhood. It appears that the existence of schizophrenia could undermine our most basic conception of the self, that only 'I' have access to, own and control my conscious mental realm or mind, and 'I' know 'my' mind directly. Schizophrenia seems to illustrate that this agential singularity is an illusion that disintegrates when things go wrong. At the same time, some schizophrenics also claim responsibility for acts that are not their own. They can be deluded and think they control the weather, traffic or even other people's actions.<sup>14</sup> This means that their self-boundaries can be both porous and diffuse. This makes the self/other distinction hazy.

It appears that schizophrenia argues against one of the more robust claims about self-knowledge, the belief in 'immunity to error through misidentification' which I have referred to in previous chapters.<sup>15</sup> It is widely held that one cannot be mistaken about certain types of self-ascription. If you use the term 'I' when referring to yourself as the subject of an experience, then you are always right about who you are referring to. Wittgenstein claimed that it is nonsensical to ask of someone (for example) who is in pain, 'are you sure it is your pain?' One can say of others that they doubt your pain but it doesn't make sense to say it about oneself.<sup>16</sup> Pain is not something we know but something we have. This is something one cannot be mistaken about in that you can't misidentify who has the pain. As already mentioned, you may be wrong about what you

---

<sup>13</sup> Daprati, et al, 1997, pp 71-86.

<sup>14</sup> See Heinrich, 2001, p. 4.

<sup>15</sup> See Shoemaker, 'Self-reference and self-awareness', *Journal of Philosophy*, Vol. 65, no.19, pp. 555-567, 1968.

<sup>16</sup> Wittgenstein does not claim there is a distinct self or subject but does distinguish between referring to 'I' as object (I am wearing a blue jumper) and 'I' as subject (I see a bird). L. Wittgenstein, *Philosophical Investigations*, s246, p. 96 in the revised edition P.M.S. Hacker and Joachim Schulte, Blackwell Publishing, U.K., 2009.

are experiencing but you won't be wrong about who is experiencing it. In the same way, one cannot be mistaken about one's experiences even if one is mistaken about what is causing them. This is because one experiences oneself as a *subject* of that experience while the experience is happening. According to Shoemaker, this knowledge is direct and unmediated. It is not inferential like other forms of knowledge.<sup>17</sup> However, this immunity to error appears to be challenged by schizophrenic delusions of control.

Numerous experiments have been conducted in an attempt to pull apart what (malfunction of) mechanisms are responsible for producing such extreme and bizarre phenomenology.<sup>18</sup> Current theories of action have separated out two different causal pathways or chains.<sup>19</sup> Mirror experiments with children demonstrate that there may be two distinct processes that contribute to self-knowledge. The experiments involved secretly placing labels onto the children's foreheads while they were playing and then showing them images of themselves on video or using a mirror to observe their reactions. Two year olds are unlikely to react to the sticker when shown the video, whereas older children (4 year olds) will attempt to remove the sticker. All the children reached for the sticker when shown a mirror.<sup>20</sup> According to Carruthers, children need to know both that the body in the mirror (or video) is theirs and that they can affect changes to it by attempting to remove the sticker. The first process is the capacity to know the boundaries of oneself, either external in relation to the limits of one's own body or internal in relation to one's own thoughts. This process contributes to one's sense of ownership, in labelling things 'mine.' The other capacity is to know that one is controlling one's actions, both internally and externally. This creates the sense of agency, i.e. that one's actions are under one's control. Carruthers terms the two processes the boundary function and the agentive function. Both functions have a bodily

---

<sup>17</sup> Shoemaker, 1968.

<sup>18</sup> See C. Frith, 'The self in action: Lessons from delusions of control,' in *Consciousness and cognition*, Vol. 114, 2005, pp 752-770.

<sup>19</sup> Glenn Carruthers, 'A Model of the Synchronic Self' in *Consciousness and Cognition*, Vol. 16, 2007, pp. 533-550.

<sup>20</sup> Carruthers claims that this shows that 4 year olds have a diachronic self, not just a synchronic self as the video is delayed feedback rather than immediate. It may also be that 4 year olds are better at drawing inferences than two year olds, such as the sticker must still be there. Does this require a separate self capacity?

aspect and a mental aspect.<sup>21</sup> He uses the evidence from the mirror experiments to claim that mirror recognition requires both a boundary function and an agentive function.

It would appear that schizophrenics can have a disorder of either mechanism. They can claim that something lies within the boundary of their self, as in a thought, but claim no sense of control over it. This has an impact on their claims of agency over their actions. Likewise, their sense of agency can extend to and include things that fall outside their boundaries, like the actions of other people or things. Much has been written about the ramifications of the results in relation to the unity of self and the veridicality of agency, mostly from a psychological perspective. Not only does the reported phenomenology of schizophrenic sufferers seem to undermine the idea of a singular agent, but it seems to cast doubt on the incontestability of first person accounts. It would argue against the claim that subjectivity is experiential and doesn't require judgement. Subjective self-ascriptions are supposedly directly experienced, hence immune to the kind of mistakes that can occur with other ascriptions. If self-claims are, in fact, inferential, they are not direct and this means we have no special access to the contents of our minds. I could no longer report that I am in pain and have it treated as an infallible statement of self-knowledge. I could be mistaken. This would cast doubt on the veridicality of our self phenomenology and, thereby, weaken any claims we make about the existence of a subjective self. I address this point in some detail in Chapter 5 when I critique both Dennett and Metzinger for dismissing phenomenal experience in this way.

A series of studies conducted on agency attribution in schizophrenics adds weight to the view that the mechanisms involved in the normal attribution process may be malfunctioning. One study by Daprati et al involved a series of experiments designed to uncover the mechanisms involved in attributing an action to its proper agent.<sup>22</sup> This required participants to perform a specified intentional action and then to record how this action was reported (as my action or another's action). The study used videos and

---

<sup>21</sup> Carruthers refers to these as the agentive<sub>B</sub> and agentive<sub>M</sub> selves and the boundary<sub>B</sub> and boundary<sub>M</sub> selves, 2007, p. 538-539.

<sup>22</sup> Daprati et al, 1997.

mirrors to mask the identity of the true agent of the expected/unexpected action to ascertain how accurately people recognised their own actions or how readily they attributed an action to themselves as the agent. In reality, someone else's hand was used to perform the same action the participant intended to perform, but by making it look as though the alien hand were their own. Sometimes the alien hand would be slightly out-of-sync with the genuine hand and sometimes it performed unexpected (unintentional) acts. The schizophrenic patients used in the experiments were those known to suffer from delusions of control and/or hallucinations. The results showed that, compared to the control group, the schizophrenic participants were much more likely to claim ownership of the actions performed by the alien hand, even when the action performed was out-of-sync or unexpected. The researchers drew the conclusion that the mechanism for claiming ownership was the same as that used for controlling and recognising one's movements as one's own. In schizophrenics of this kind, the mechanism was considered to be deficient. This led them to falsely attribute others' actions to themselves.

Gallagher discusses the implications of similar experiments in his paper.<sup>23</sup> Like Carruthers, he draws a distinction between two different aspects of attributing an action to oneself. He claims that normally-perceiving subjects can discriminate between voluntary and involuntary movements, even though we experience both of them as residing within our own body. This indicates that there is a phenomenological distinction between our sense of ownership, the experience of something being within one's self-boundary/body, and agency, the sense of controlling one's own actions. Involuntary actions are not considered under our control but are still considered under our ownership, as in occurring in and to our own body. Those schizophrenic sufferers who appear to have problems correctly attributing agency to their own movements also have problems attributing their speech acts to themselves. They can claim either that they are forced to act in certain ways or that they are forced to say things without their intention. They still claim ownership of the actions in that it was their body (mouth) that enacted

---

<sup>23</sup> Shaun Gallagher, 'Self reference and schizophrenia: a cognitive model of immunity to error through misidentification,' in *Exploring the Self: philosophical and psychological perspectives on self experience*, ed. Dan Zahavi, John Benjamins Publishing Company, The Netherlands, 2000.

but they deny agency. Sometimes the agent is not some other being but the body part itself that is moving. (I.e. my hand is combing my hair but without my volition).

One explanation discussed by both Gallagher and Carruthers is a malfunction of the feedback/feedforward model of action that is variously described by those working in the field.<sup>24</sup> Basically the action model suggests that the intention to act creates two representational pathways, the motor controller to enable enaction and the intention-to-act map that predicts one's future movement in line with the action to be performed. In normal circumstances, the realised action would match the predicted model of the action, hence the sense that one's action is in line with what is expected. If not exactly as expected, the actor /agent would adjust their action to conform to the predicted action, thereby making it clear that the agent is in control of its action. The intention to pick up the cup is matched by the correct movement. It is mooted that, with schizophrenia, something goes wrong with the link between the motor command and the realisation of the action. One explanation suggests the malfunction has to be with the predictor or comparator mechanism; i.e. there is no intention-to-act map or the intention-to-act map is faulty. However, the act is not described as involuntary, which one assumes would be the case if there were no intention-to-act map. Carruthers suggests that if there is no representation of the *comparison* between the predicted movement and the realised movement, then the sense of agency is not present so it would feel like someone else is controlling one's actions. It is difficult to imagine an experimental situation that could discriminate enough to determine the accuracy of either explanation.

Either version seems to offer a possible explanation of why some physical actions performed by schizophrenic sufferers do appear outside their control. It is not clear why this would cause them to extend ownership to actions outside their boundaries. Other experiments indicate that schizophrenic sufferers rely on visual feedback more than proprioceptive feedback when claiming agency or ownership of actions, compared to the

---

<sup>24</sup> This model is attributed to Frith. See 'The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action,' *Psychological Medicine*, vol 17, no. 3, 1987 and 'Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action,' *Psychological Medicine*, vol 19, no. 2, 1989. See Carruthers, 2007 and Gallagher, 2000 for a detailed discussion of the motor controller/action representation model.



control groups.<sup>25</sup> If this is the case, then one might be likely to more readily claim ownership of actions that one witnesses or that happen around one, in the absence of the proprioceptive sensations that normally accompany one's own movement.

Frith claims that the problem is one of self-monitoring which arises in the motor control system. Schizophrenic sufferers with delusions of control do not have a problem controlling their movements, even if they lose visual contact. Rather they seem to have a problem with their awareness of their intentions to control their movements.<sup>26</sup> To some extent, they do not expect them, even when they are directed to act. It seems that this loss of a sense of control or agency over one's own acts can extend to claiming outside acts as one's own, if reliant on visual information only. Frith claims that the sense of agency is distinct from the sense of control. Agency is much broader and encompasses goal-directed behaviour. Non-schizophrenic participants in a study involving intentional goal-directed movements and just intentional movements showed that people are more likely to claim agency over a goal-directed movement that was intended, even if they did not perform it. These acts were more reliant on the success of the goal than on proprioceptive feedback as evidence of agency. This might explain why schizophrenic sufferers are more likely to claim agency over acts that they did not perform, but which they might think they intended. As Frith claims, they often see agency where there is none, even in random events such as the weather.<sup>27</sup> I am not sure why they allocate agency where there is none, especially given their own reduced sense of agency. This may be an unrelated consequence of their complex pathology. A common symptom of schizophrenia does appear to be an excessive imparting of significance or meaning to often random, unrelated events.

However, schizophrenic sufferers also have symptomatic hallucinations of hearing voices or having thoughts that they claim are not their own and, consequently, that they claim are not under their control. It is hard to imagine how the model above could explain these phenomena. Carruthers does suggest that vocal acts could fit the action

---

<sup>25</sup> Frith, 2005, pp. 758-59.

<sup>26</sup> Frith, 2005, p. 756.

<sup>27</sup> Frith, 2005, p. 766.

model described if one treats thought as a type of mental speech. Speech rehearsal is like action rehearsal and would entail the prediction of movement, i.e. like the larynx. Thus, one could seemingly hear internal speech by predicting the auditory signals one would get if one verbalised the imagined speech act. For schizophrenic sufferers to deny ownership or agency, there would need to be a malfunction between the predictor and the actualisation, possibly in the representation of the comparison as suggested before.

I am not sure this explanation works. First, not all thinking is rehearsed speech and not all mental speech is auditory or vocalised. If, in this instance, the schizophrenic sufferer can form a representation ‘of the predicted auditory consequences of a movement’ because they experience a thought, but there is no realisation to compare it to, then it is not clear what the malfunctioning comparison would be a representation of.<sup>28</sup> In addition, as Gallagher suggests, it does not make sense to suggest that there is a thought precursor to every thought, the lack of which will lead to a loss of agency in regards to one’s thoughts.<sup>29</sup> This seems highly implausible. First, the precursor has to be non-conscious or it will generate an infinite regression. This non-conscious precursor is, supposedly, the intention to think something (a thought-generator mechanism) that then, using Frith’s model, produces a non-conscious efferent copy that ‘generates an awareness of effort.’<sup>30</sup> An absence of awareness of effort could mean that the thinking would appear not one’s own. However, as Gallagher rightly points out, most thinking is not really directed or intentional in this way. Thus, it would be difficult to imagine a misfit between intention and realised thought such that it led to a belief that the thought was inserted. The same is true of the notion of ‘effort.’ Part of the phenomenology of thinking is its randomness and the lack of control over what we think about and where the thoughts go. This ‘unbiddenness’ can be stressful when the thoughts are unpleasant or even obsessively repetitive about one event, but this doesn’t normally lead us to a belief in alien thought insertion. For this to occur, there has to be something extremely odd about the way schizophrenic sufferers experience the world and their own mental states that could create such bizarre illusions.

---

<sup>28</sup> Carruthers, 2007, pp. 546-547.

<sup>29</sup> Gallagher, 2000, pp. 215-217.

<sup>30</sup> Gallagher, 2000, p. 215.

While it is possible to proffer an explanation of what may be going wrong in schizophrenics, it also needs to match up with (if not explain) the phenomenology. Gallagher puts forward what he claims is an alternative explanation, drawing from Husserl. He claims that a better model is one using the concepts of protention and retention which Husserl used to explain time-consciousness. If one imagines perceiving an enduring object, like listening to a piece of music, one is first conscious of the first note which is present in one's current state of consciousness (now-phase). As one continues to listen, each new now-phase also contains a just-past phase which is a *retentioning* of the previous now-phase and its previous just-past phase. This creates a retentional continuum. At the same time, there is a simultaneous anticipation of the about-to-happen experience, or *protention*. One's consciousness is now directed towards a 'longitudinal intentionality' that provides a sense of temporal continuity to both the perceived object and the perceiving consciousness.<sup>31</sup> With our own actions, including thought and speech, we have a sense of intended direction or where we intend to go (protention). Gallagher claims that retention includes a 'double intentionality' in that one is both aware of the object of experience and aware of oneself undergoing the experience. Thus, one is non-reflectively aware of 'me' existing in time.

He then argues that protention would have a similar double intentionality because one both anticipates the next note/next act and also that it will be 'me' hearing or doing it. This sense, he claims, provides a sense of agency and ownership that is not retrospective but built into the process itself. Gallagher claims that the retention-protention structure is already built into consciousness because of the temporal nature of consciousness. I can (normally) track my thoughts and predict where they might go. Thus ownership is already built into the stream of consciousness, making us claim the thoughts are within our own minds.<sup>32</sup> For schizophrenic sufferers, there could be a breakdown in the protentional function. This could cause the lack of a sense of generating one's thoughts because this is the mechanism that tracks the thoughts and predicts their direction. If one has no idea where one's thoughts came from or where they are going, then it is likely

---

<sup>31</sup> Gallagher, 2000, pp. 222-224.

<sup>32</sup> Gallagher, 2000, p.229.

they would seem to be outside one's control. This would make their direction surprising. It is not clear, however, why they would seem not one's own.

#### **4.4 Assessing the credibility of current explanations of the mechanisms of schizophrenia**

I am not sure that the models suggested above do the work they claim to. As Gallagher points out, Frith's model is more complicated than it needs to be and does not deal well with the key symptoms of schizophrenia, such as alien voices and thought insertion. While Carruthers attempts to address this with a version of Frith's model, his version only works by making all thoughts linguistic and potentially involving motion. This seems an ad hoc solution. Neither model addresses what Gallagher calls the 'occasionality' problem.<sup>33</sup> Schizophrenic sufferers of the kind discussed here do not experience alien thought insertion all the time. They may have long periods where the voices go away entirely and other periods where they are present more often. This symptomatic phenomenology is not representative of the rest of their thinking or their experience of their mental realm. Outside of the delusional periods, their mental world would appear to be like non-schizophrenics. If deficits like those described above are responsible for the delusions, then it is not clear why they only operate occasionally. Although there may be an explanation for this, it is not covered within either Frith's or Carruthers's models.

At the same time, only certain kinds of thoughts are being inserted, some of the time. The alien thoughts or voices tend to have a repetitive theme or content or phenomenology (i.e. the same tone of voice, mood, similar content or similar identity of the person inserting the thoughts). Sometimes, they come from a consistently named person (e.g. someone called Chris<sup>34</sup>). This doesn't fit a generalised malfunction model. I would also claim that the occasionality problem is a problem for agency of actions as well. Not every action performed by someone with schizophrenia is designated as outside their control. In fact, the vast majority of their actions are owned and acknowledged as theirs. Neither Frith's model nor Carruthers's amendments to it explain

---

<sup>33</sup> Gallagher, 2000, p. 222,

<sup>34</sup> Heinrich, 2001, p. 5.

why schizophrenics are only deluded about some actions and not others, and why this happens some of the time rather than all of the time. Semantic content, as indicated by the content of the thought or the nature of the action, could play a role here but what content and why needs to be explained.

Unfortunately, Gallagher's retention/protection model does not seem to fair any better. He introduces a temporal element to accommodate the effect that certain intentions or situations have on our experience of time. He claims that one's non-conscious intentions could cause a desynchronisation between 'the phenomenological stream and the stream of efferent copy.'<sup>35</sup> The experience of loss of control or agency is experienced as a kind of negative or counter 'déjà vu,' because one stream of information lags behind the other. Hence, the thought just 'pops up.' However, there are many other situations when one's thoughts can be completely outside one's control, intrusive or take one by surprise, depending on stress levels and traumatic past experiences. Post traumatic stress disorder is a case in point.<sup>36</sup> In non-psychotic, everyday cases, thoughts can just 'pop up' unintended and with surprising content. This does not lead to bizarre claims of alien insertion or lack of agency. Gallagher does suggest that the occasionality of the delusions may be the result of situational factors like stress or anxiety that occur only at certain times. Again, while it is likely that schizophrenic sufferers are more likely to have psychotic episodes under certain environmental conditions, there is no evidence of a pattern of events that trigger episodes. One would still assume that if relevant perceptual feedback/feedforward systems were malfunctioning, the delusions of control would be more persistent, more predictable and more global.

Some schizophrenics claim that the experience they have is as though someone else knows what they are thinking or can pre-empt what they are going to think (sometimes called 'thought broadcast.').<sup>37</sup> This implies access to pre-knowledge of one's thoughts

---

<sup>35</sup> Gallagher, 2000, p. 221.

<sup>36</sup> Scott O. Lilienfeld, Steven Jay Lynn, Irving Kirsch, John F. Chaves, Theodore R. Sarbin, George K. Ganaway, and Russell A. Powell, 'Dissociative Identity Disorder and the Sociocognitive Model: Recalling the Lessons of the Past', in *Psychological Bulletin* 1999, Vol. 125, No. 5, 507-523.

<sup>37</sup> See Gallagher, 'Self reference and schizophrenia: a cognitive model of immunity to error through misidentification,' 2000.

and could indicate that the protentional mechanism or Frith's feedforward system is, in fact, working. In this instance, the difference appears to be that, unlike our normal thinking process, the schizophrenic is aware or conscious of the protentional/feedforward mechanism such that they experience a kind of third person déjà vu. The retention-protention mechanism could be desynchronised as Gallagher suggests but it results in a different phenomenology to that of thought insertion. Interestingly, as Gallagher describes it, schizophrenics appear to have overly complicated thinking processes where they appear to be aware of stages in their thinking processes that are normally not available to consciousness. If this is the case, it would change their phenomenology such that their actions or thoughts may not appear to be under their control. The controversial findings of Libet in the early Eighties indicated that this is actually the case for all of us.<sup>38</sup> If one were suddenly aware of a thinking process usually not accessible to consciousness, and the concomitant phenomenology, such as an affect or engagement with the thought predictor, was not present, it could seem as if someone else were tracking one's thoughts.

It is worth noting that schizophrenic sufferers don't appear to have the same sense of affect or ego engagement that is normally associated with one's own actions or inner world. The 'what it is like' or qualia attached to the processes of action only seem to come into play at a particular point in the thinking or acting process. Thus, the self-conscious experience of schizophrenic sufferers appears to be quite different from those of non-schizophrenics. This relates back to a flattening of affect, as set out in the DSM IV symptoms. Frith suggests that the 'normal' suppression of the proprioceptive response, evident when we are engaged in goal-directed actions, is not suppressed in schizophrenics, making it more likely they will claim an action to be theirs when it is not.<sup>39</sup> If there isn't the transparency in the accompanying phenomenology, if one's normally non-conscious bodily or cognitive processes are experienced instead of

---

<sup>38</sup> B. Libet, C. A. Gleason, E. W. Wright, and D. K Pearl, 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act,' *Brain*, Vol. 106, 1983, pp 623-642.

<sup>39</sup> See Frith, 2005, pp.764-765. However, in other instances it seems like the proprioceptive response is subdued, so this is not a consistent factor.

masked, then the resulting actions may not seem like one's own.<sup>40</sup> There will be some additional steps in the cognitive process that are normally inaccessible to our awareness. The mere occurrence of the thought or intention inside the head would not be enough to override the (delusional) belief that the ideas and internal voices were not their own. Schizophrenics don't claim ownership; they *disown* the thoughts. A sense of ownership must come from more than just physical occurrence inside the same body or within the body's boundaries pace Carruthers. It seems that the kind of afference that makes random, spontaneous or unbidden thinking feel like it is still our own is missing in schizophrenic sufferers. It would seem that schizophrenic sufferers lack the requisite self-phenomenology, what Varela calls affective tonality.<sup>41</sup> This is not accounted for in either Frith's, Carruther's or Gallagher's model. The issue of why it affects some thoughts rather than others is still a problem, however.

#### **4.5 Implications for the self and sense of self**

It has been suggested that schizophrenic delusions of control and agency argue against the existence of a singular or unified self. In particular, they seem to provide evidence against the immunity principle. Self-conscious thought presupposes a capacity to think about oneself and a capacity to know certain things about oneself. It involves the unique capacity to apply concepts and descriptions to ourselves—what is called the capacity for 'I' thoughts. 'I' thoughts involve a distinctive type of self-reference; it is not just the capacity to refer to ourselves as we do to all objects as an identification of that object (here is a chair; here is Sandra), but it is the use of what Perry called the 'essential indexical.'<sup>42</sup> Even if I forget my name, where I am, etc., I can still refer to myself as me. I know who I am as a specific instance of the word 'I'. As such, I am never not recognisable or identifiable to myself and I cannot be mistaken in my use of that indexical. This is the capacity for 'identification-free self-reference' and the capacity of

---

<sup>40</sup> Metzinger claims that the transparency constraint is essential in creating our phenomenology as it masks the underlying non-conscious cognitive processes. See Chapter 3, *Being No-one*, 2003.

<sup>41</sup> F. Varela cited in S. Gallagher, 2000.

<sup>42</sup> John Perry, The Problem of the Essential Indexical. *Nous* Vol. 13, no1, 1979, p. 3-21.

'immunity to error through misidentification.'<sup>43</sup> One cannot think an 'I' thought without knowing that it is about oneself and one doesn't need to refer to anything else to identify oneself to oneself.<sup>44</sup> It is argued that schizophrenic sufferers lack this irrefutable self-knowledge.

There do appear to be issues with ownership and correct attribution of thoughts and actions in some schizophrenic sufferers. A schizophrenic patient could say 'there is a thought in my head but I am not thinking it,' or 'there is a voice in my head and it is not mine.' Whether this is due to a malfunction in the representational system or a lack of affect is unclear. Either way, it would be difficult for a non-schizophrenic person to imagine coming to a similar conclusion. If one is aware of a thought it is unlikely one would attribute it to someone else. On the other hand, it would be false to claim that schizophrenics do not use 'I' appropriately or that they use it mistakenly. When they do claim ownership, it is immune to error in the same way as non-schizophrenics. Neither do they claim ownership of other people's thoughts. While they do sometimes claim ownership or agency over acts that are not their own, this is usually at a distance (i.e. a change in the weather) or involving judgements that rely on visual feedback. The experiment conducted by Daprati et al with the 'alien' hand, was a purely visual discrimination test, focusing on the role that visual input has in belief formation of this kind. In these experiments, the self is viewed as object, not subject. Both the control group and the schizophrenic group made errors of judgement in act attribution, although this was less likely to be the case in the control group when the alien arm performed unpredictable actions. The experiment did not take into account the subjective experience of performing an action, i.e. the sense of embodied action or the proprioceptive feedback that makes a person claim it was their arm rather than someone else's. In itself, it does not challenge the immunity to error principle. It does, however, support the idea that imputing self or other agency in goal-directed actions is largely inferred rather than relying solely on directly experienced proprioceptive feedback. It

---

<sup>43</sup> See Carol Rovane for a discussion of these terms and why the use of 'I' always refers; 'The Epistemology of First-Person Reference,' *Journal of Philosophy*, 1987, vol. LXXIV, no. 2, pp. 147-167.

<sup>44</sup> This runs counter to Hume's claims about epistemology and the self. See Hume, Bk I, part IV, s VI, 'Of Personal Identity' in *A Treatise of Human Nature*.



supports the idea that it is easier to be tricked when addressing the self indirectly as object than it is by experiencing the self as subject.

Does the exclusion of certain self phenomena from the ‘I’ category violate the ‘immunity principle?’ Are there grounds to claim that the use of the indexical is inferred in the same way knowledge of other things is inferred? We know that schizophrenics generally perceive their own thoughts just as others do, and yet they also perceive some thoughts that they claim are not their own. So something is different about the alien voices; they have some phenomenal quality that leads to a bizarre claim. One could counter claim that all the schizophrenic sufferer is doing is *correctly* reporting on their experience of what it is like to be them. They say ‘I hear voices in my head,’ and they do. From their perspective, this is the case. Their reports about themselves are still immune to error. The ‘delusions,’ on the other hand, do appear to be inferential. I am not convinced that this casts doubt on our capacity to make self-knowledge claims. Schizophrenic sufferers are, in fact, right that something is wrong or amiss; their phenomenology has dramatically changed. This leads them to draw bizarre inferences in an attempt to explain the changes, such as they now share their mental space, albeit sporadically, with some other entity.<sup>45</sup> I am not convinced that this phenomenon is evidence that the self is not intrinsically singular or that it does not exist.<sup>46</sup>

Carruthers claims that, ‘the self... is the set of cognitive capacities that underlie the various senses of self.’<sup>47</sup> In his account, these are the boundary<sub>B</sub> and boundary<sub>M</sub> capacities and the agentive<sub>B</sub> and agentive<sub>M</sub> capacities, which he refers to as ‘selves.’<sup>48</sup> He then surmises that, if these can be dissociated from each other as appears to be the case in schizophrenia, then by implication there is no self, that it has been explained away. While I am sympathetic to many of his claims about self-functions, this conclusion seems unwarranted, a case of applying one’s ontological beliefs to some

---

<sup>45</sup> Schizophrenics have similar delusions in that the voice is always negative, nasty and demanding they perform retributive acts. This is sometimes accompanied by visual delusions that can lead to mistaking others as evil or as impediments. Heinrich 2001.

<sup>46</sup> There are other ailments that can fragment the sense of self that I address later in this chapter.

<sup>47</sup> Carruthers, 2007, p.548.

<sup>48</sup> The boundary and agentive capacities are further divided into mind and body domains.

things rather than others. As with so many writers in this field, it appears that the holy grail of finding a singular Cartesian self-neuron still holds sway.<sup>49</sup> Either a ‘real’ self will have *no* neural mechanisms underpinning it, which would make it some ineffable Cartesian ego, or should we find some underlying mechanism or mechanisms that take the role of self, then this would not count as a self. I address this odd reductive position in the next chapter. There are plenty of examples of composite and complex objects in biology. The visual system, for example, is constituted by multiple functional units but this does not lead us to conclude that there is no visual system or no vision, even though we may argue it is not quite what we think. It just means that the visual system is more complex than was first thought. Ditto the self. If any mechanisms are found to underpin the self they are bound to be complex and complicated. Why should the existence of mechanisms supporting the sense of self lead to a sceptical conclusion in regards to the existence of the self? If anything, they add weight to the idea that the sense of self is biologically grounded, rather than being solely a product of a self-narrative. By all accounts, schizophrenic sufferers still have a sense of self; they still have a strong sense of ‘I.’ Interestingly, there is still a single subjectivity or consciousness. The so-called alien presence is very narrowly constrained; it can’t be accessed from a first-person perspective. In itself, this phenomenon does not argue against a naturally unified or singular self. However, it does indicate that the phenomenon of the self can be damaged.

There are reasons why aspects of our self-phenomenology start to fall apart. While the aetiology of schizophrenia is noted for its heterogeneity, there is evidence that there are some common underlying neurological abnormalities present.<sup>50</sup> For instance, Frith says schizophrenic sufferers show abnormalities such as hyperactivity in the parietal cortex, common in patients with delusions of control.<sup>51</sup> Feinberg indicates that neuropathologies of the self, such as schizophrenia and Capgras (often found in those with schizophrenia) demonstrate lesions to the frontal and medial-frontal regions of the brain. He hypothesises that the right frontal zones play an integral part in setting the ego

---

<sup>49</sup> Of course, if we were to find such a thing, then it would have to count as a Cartesian self.

<sup>50</sup> See M. Coltheart, R. Langdon and R. McKay, ‘Schizophrenia and monothematic delusions,’ in *Schizophrenia Bulletin*, Vol. 33, no. 3, 2007, pp 642-647 and Heinrich 2001, for a discussion on the complex etiology of schizophrenia.

<sup>51</sup> Frith, 2005, p.766.

boundaries of the self.<sup>52</sup> Such findings seem to indicate that the delusions, if not their semantic content, are caused by a malfunction in the mechanisms responsible for producing a sense of self-boundedness. The damage seems to put some sensations or thoughts outside of the self-boundary, normally produced by specific mechanisms in the brain that enable the system to individuate itself such that it is represented as singular. This does not lead to the conclusion that there is no self or that the self is not intrinsically unified or singular. What it indicates is that the sense of self may have a strong neurological basis and that this can be damaged in such a way as to produce a changed sense of self or a damaged sense of self. This tells us something about the underlying mechanisms responsible for the sense of self and about what we might consider as a candidate for the self or self-system.

#### **4.6 Dissociative Identity Disorder (DID)**

There is another neuropathology of the self, now referred to as a type of dissociative identity disorder, (previously known as Multiple Personality Disorder or MPD) that appears to raise even more concerns about the singularity of the self than schizophrenia.<sup>53</sup> The possible existence of multiple personalities or selves in the one body or brain challenges our every day conceptions of what it is to be a thinking conscious person or self.<sup>54</sup> As singular people, it is hard to imagine how there can be one person or personality acting through a body one minute, and another completely different person or personality acting through the same body the next. This seems even stranger if we imagine that each currently operating identity or self may be unaware not only of the other's existence but also of their actions while they are 'hosting' the shared body. In the case of Sybil (a well-documented case of MPD made famous in film), there were supposedly up to sixteen distinguishable personalities operating through the one

---

<sup>52</sup> T. E. Feinberg, Neuropathologies of the self,' *Consciousness and Cognition*, Vol 20, 2011, pp.75–81

<sup>53</sup> The American Diagnostic Statistic Manual officially changed its classification of multiple personality disorder (MPD) in the 1994 version (DSMIV) to dissociative identity disorder (DID) see <http://www.didmpdinfo.com/DSMIV.html> .

<sup>54</sup> R. B. Allison was the original psychiatrist to claim MPD was a dissociative disorder. He argues strongly against its name-change to DID as he claims this denies the reality of multiples. See Allison, 'On discovering multiplicity', *Svensk Tidiskrift fur Hypnos*, 1978, Vol 2, pp. 4-8.

body.<sup>55</sup> Each one appeared to be self-conscious and capable of verbalising its own specific needs and interests, and each one displayed its own unique personality. According to some reports, there can be hundreds or even thousands of different ‘selves’ in the same brain of a patient diagnosed with DID (although most of these vanish under therapy).<sup>56</sup>

If ‘multiples’ really do exist, they appear to be the paradigmatic counterexample for the existence of a same continuing self, particularly if they can also be co-existent or co-conscious with at least one other self or ‘alter.’ The possibility of multiple personalities raises several issues in relation to any conception of the self. First, it argues against the idea of intrinsic singularity. Second, the lack of shared consciousness seems to argue against any inherent unity of self. Third, the serial nature of their appearance and the diversity of their personalities seem to indicate multiple agency, i.e. that there is more than one agent acting out of the one body and each one has its own needs and desires. If this is genuinely the case, then it would appear to support either a Humean conception of the self or a model like Dennett’s multiple drafts.<sup>57</sup> Dennett has argued that what the ‘multiples’ demonstrate is actually what is true about ourselves as complex cognitive systems but which is generally hidden through a process of socialisation. He claims there is no inner self, no single agent operating in the world. There are just mixed bundles or ‘multiple drafts’ of experiences that randomly cohere around a point in time to create the illusion of a distinct but transitory personality. That personality would not be reflective of any ‘real’ self and could develop into any type of person with any variety of traits. Oppenheimer shares a similar view.<sup>58</sup> He argues that normal development consists in the integration of sub-personalities into one self-system. Thus,

---

<sup>55</sup> The film ‘Sybil’ was based on the case of Shirley Ardell Mason. A Dr Rieber now claims that tapes he has confirm that Mason’s doctor, Cornelia Wilbur and Fiona Shreiber, the author of the book, exaggerated her symptoms. This has thrown doubt on the reality of Mason as a ‘multiple.’ See ‘Tapes raise fresh doubts on "Sybil" case in U.S.’, San Francisco, *Reuters News*, August 17<sup>th</sup>, 1998.

<sup>56</sup> See August Piper, and Harold Merskey, ‘The Persistence of Folly: A Critical Examination of Dissociative Identity Disorder. Part I. The Excesses of an Improbable Concept,’ *Can J Psychiatry*, 2004, Vol. 49, pp. 592–600.

<sup>57</sup> Dennett, 1991, pp. 111-115.

<sup>58</sup> L. Oppenheimer, ‘Self or selves? - dissociative identity disorder and complexity of the self-system,’ *Theory & Psychology*, Vol 12, no.1, 2002, pp. 97-128.

DID is a ‘naturalistically-occurring’ condition because that process of integration failed to occur. DID is proof positive that there is no intrinsic inner self.

#### **4.7 Life as a multiple**

In order to investigate DID and its potential challenge for my conception of the self, we need to get a clearer picture of what it is like to be a multiple. What is the phenomenology? There are several currently existing websites that have been (supposedly) set up by multiples to disseminate information about themselves, to advocate for their right to exist as multiples and to lobby for their acceptance in society as an alternative way of being, in comparison to the more traditional ‘singlets’ way of being.<sup>59</sup> From these sites, we can get some idea of how they see themselves. First, multiples do not consider themselves to be suffering from a dissociative identity disorder. As far as they are concerned they really are multiple systems, regardless of their aetiology. They do not consider themselves delusional. Second, they talk about themselves as though they were multiple people living in a shared house. If one accepts their claims of what it is like, then there exists a number of host bodies that house a collection of completely separate, easily distinguishable, robust, co-existing entities. These entities or alters can consist of either gender, are often other-handed, of various ages, and with different memories, skills, tastes and personalities to each other. According to one account, ‘Each of us has our own hopes and dreams, our own range of feelings, our own ideals.’<sup>60</sup>

In addition, there are some multiples who claim that they have always been multiple and that they have no history of severe trauma or abuse. They claim no loss of memory and say they have had a long association with their alters since childhood.<sup>61</sup>

---

<sup>59</sup> See *Pavillion: voices of plurality in action*, <http://www.karitas.net/pavilion/>, accessed May, 2010, or Anthony Temple, *This is our truth*, 2007, <http://astraeasweb.net/plural/ourtruth.html>, accessed May 2010.

<sup>60</sup> See V of the Anachronic Army, *Why We Do Not Have MPD/DID*, <http://www.astraeasweb.net/plural/v.html>

<sup>61</sup> Temple claims that he was not the victim of trauma and has always been aware of his other selves without attending therapy – ‘The third criterion [DSMIV] completely disregards the fact that many multiples operate with *no* loss of memory (or, more properly, continuity of consciousness between upfront personnel). They possess subjective libraries or common memory pools that allow anyone who comes up

Temple, a self-professed multiple, claims that:

...there are people who are living sane, functional, healthy, productive lives as multiples; undiagnosed, not in therapy, but knowing they are many. Not broken pieces of a single, original person, but many people sharing the same body. Not made-up hysterical delusions, but responsible individuals who cooperate amongst each other and support each other just like a family. Not seeking to blame mistakes and bad behavior on 'alters,' [they] live by codes of group responsibility. 62

However, not all multiples agree with each other. While they all appear to accept the reality of being a multiple, as opposed to being deluded, and that they are a multiple *now* regardless of what they were before, they do not all agree that it is an acceptable alternative way of being. There are other multiples who believe that it is a disorder. They agree with the diagnostic clinicians that it is only through therapy that they can become aware of their condition and that it is only through therapy that they can be (re-) integrated.<sup>63</sup> They claim that multiples who avoid therapy are in denial.<sup>64</sup> For them, integration into a *single* identity is the only desirable outcome.

These views capture what is at issue. Some argue that the multiple state is natural and represents an alternative, equally legitimate way of being. Others argue that the multiple state is an anomaly brought about by trauma and that the natural state is to be singular. This second position seems to be claiming that there is one self but with multiple identities.

#### **4.8 How real are multiples and DID?**

The disagreement amongst the multiples themselves about exactly what is going on highlights an important issue. The very existence of multiples is itself a highly controversial and hotly debated subject within psychology and psychiatry. There is a

---

front to be in immediate command of a situation...and to have full knowledge of the past. <http://astraeasweb.net/plural/ourtruth.html>; accessed May 2010.

<sup>62</sup> <http://www.karitas.net/pavilion/> accessed June, 2010.

<sup>63</sup> Anthony Temple, <http://astraeasweb.net/plural/ourtruth.html>

<sup>64</sup> Anthony Temple also puts the opposite case; "People who are truly multiple cannot possibly know that they are. Even in a group with many frontrunners, none of them can know each other or have any sort of communication. In real multiple systems, the frontrunner never knows of the others until the diagnosis is made by a qualified therapist."

great deal of disagreement between clinicians, therapists and researchers about whether the condition is a genuine disorder or not. Some believe quite strongly that it is either iatrogenic, actually created within the therapeutic environment by the therapist, or it is an (albeit subconscious) artifice of the patient to gain attention.<sup>65</sup> Of the therapists and researchers who do accept it as a genuine disorder, there is disagreement about its characterisation. Some feel the original focus on multiple personalities was more accurate, while others support its re-description as a dissociative disorder.<sup>66</sup> Within the former framework, there really are multiple personalities that have been created by the patient's unique past; within the latter, the patient is treated as delusional and their other personalities are considered to be delusions similar to those manifested by schizophrenic sufferers.

The change from MPD to DID occurred in the 1994 version of DSMIV. Section 300.14 sets out four necessary diagnostic conditions for DID. These are:

- The presence of two or more distinct identities or personality states;
- At least two personalities must take control of the person's identity on a regular basis;
- Exhibits aspects of amnesia, that is, the person forgets routine personal information;
- The condition must not have been caused by "direct physiological effects," such as drug abuse or head trauma.<sup>67</sup>

It is widely accepted that experience of extreme and prolonged trauma in early childhood is at the heart of the condition.<sup>68</sup> While there may have been a contributing traumatic event such as witnessing the (violent) death of a parent or sibling, the most common trauma is either extreme physical or sexual abuse. Although there are

---

<sup>65</sup> Spanos and his followers have advocated for this position for the last 35 years. N. P. Spanos, 'Multiple identity enactments and multiple personality disorder: A sociocognitive perspective', *Psychological Bulletin*, Vol.116, 1994, 143-165.

<sup>66</sup> See Ian Hacking, *Rewriting the Soul: multiple personality and the sciences of memory*, Princeton University Press, Princeton, 1995, and Dr Ralph Allison for an alternate personal account.

<sup>67</sup> Dissociative Disorders from Diagnostic Statistic Manual IV; <http://www.didmpdinfo.com/DSMIV.html>.

<sup>68</sup> A typical entry for DID – 'Dissociative identity disorder (previously known as multiple personality disorder) is a fairly common effect of severe trauma during early childhood, usually extreme, repetitive physical, sexual, and/or emotional abuse.' <http://www.webmd.com/mental-health/dissociative-identity-disorder-multiple-personality-disorder>

exceptions, the abuse has to be severe, it has to be prolonged and it must begin in very early childhood. As the most common form of abuse is sexual, patients diagnosed with DID tend to be predominantly female. The current theory within psychology is that DID is a coping mechanism to preserve the core self; the experiences are so extreme that the 'ego' brackets them off into another realm so that the central core self remains unaffected.<sup>69</sup> There is no apparent memory of the trauma by the presenting person or personality, and they have no apparent memory or knowledge of the 'alter egos' that are subsequently created.

The difficulty for clinicians in diagnosing DID is that patients must satisfy the diagnostic criteria set out in DSMIV. However, the first two (defining) conditions only become apparent under hypnosis. The defence for using hypnosis as part of the therapeutic treatment is that the patient suffers from amnesia, the third diagnostic condition for having DID, and hence will have no memory of any childhood abuse or of the 'alters' that may exist. In the majority of cases, alters emerge for the first time during therapy, usually under hypnosis.<sup>70</sup> Commonly, it is during therapy that the patient with DID becomes aware of their alters for the first time. There is some evidence to suggest that the number of alters increases as therapy continues, often reaching the hundreds.<sup>71</sup> Tales of physical or sexual abuse also emerge during therapy and these stories appear to increase in severity as hypnosis therapy continues.

In some cases, patients begin to report horrific memories of satanic rituals, rape, murder and even cannibalism.<sup>72</sup> Neither the therapist nor the patient would have been aware of these dramatic events if it were not for the therapeutic environment and the use of hypnosis. Such a 'chicken and egg' situation lends itself to accusations of manipulation by the therapist, or of pretence or confabulation on the part of the patient, even if subconscious.

---

<sup>69</sup>Dissociative disorders, DSM IV.

<sup>70</sup> Paulette Gillig, 'Dissociative Identity Disorder: A Controversial Diagnosis' in *Psychiatry* (Edgemont) 2009;6 (3):24–29.

<sup>71</sup> Klufft RP. Dealing with alters: a pragmatic clinical perspective *Psychiatr Clin North Am* 2006; 29 (1):281–304.

<sup>72</sup> Lilienfeld et al, 1999, pp. 507-523.



There is a vast body of literature that is sceptical of the claims made by clinicians about the existence of DID.<sup>73</sup> After a comprehensive analysis of the literature, Piper, a psychologist, came to the following conclusions:<sup>74</sup>

- the contemporary diagnostic criteria are vague and over-inclusive;
- the recent alleged increase in prevalence of the disorder is almost certainly artefactual;
- there is little literature support for the theory that DID/MPD results from childhood trauma;
- many of the techniques used to diagnose and treat the condition reinforce its symptoms.

There do seem to be some doubts about its aetiology. Even though the currently accepted cause of DID is extreme and prolonged trauma, (generally violent sexual abuse in early childhood), there are some people diagnosed with DID who have no history of trauma (about 30%).<sup>75</sup> One would also assume that trauma which is extreme enough to cause such spectacular effects would leave some social traces, such as hospital records, welfare reports or other witness accounts. In reality, there is little independent corroboration of extreme abuse, including evidence of physical damage such as scarring, broken bones or even pregnancies. Where there is documentary evidence of extreme abuse in some well-publicised cases of known long-term abuse of children, as in cases of family psychopathy, or kidnap and sexual slavery, there has been no parallel evidence of DID presenting in the surviving victims.<sup>76</sup> In addition, there is no reported evidence of multiples presenting in young children or teenagers, which is supposedly where the fracturing of ‘self’ begins.<sup>77</sup> The multiple selves or ‘alters’ do not appear until adulthood

---

<sup>73</sup> See e.g. Lillienfeld et al, 1999, and August Piper and Harold Merskey, ‘The Persistence of Folly: A Critical Examination of Dissociative Identity Disorder. Part I. The Excesses of an Improbable Concept,’ in *Can J Psychiatry*, Vol. 49, 2004, pp. 592–600 for two comprehensive literature reviews.

<sup>74</sup> A. Piper jr., ‘Multiple personality disorder’, *The British Journal of Psychiatry*, Vol. 164, 1994, pp. 600-612.

<sup>75</sup> See Piper and Merskey, 1994 and Hacking, 1995.

<sup>76</sup> Oppenheimer, 2002, p.114. Horrors of this kind often result in death. However, there have been some reported cases where the victims have survived. There have also been several reported cases where children have been removed from their home due to prolonged torture, sexual abuse and neglect by their parents.

<sup>77</sup> Alters may be manifested in young children as ‘imaginary friends’ but, given how common this occurrence is in most children, it would be hard to draw that conclusion with any confidence.

and most manifest only after therapy. Police who questioned a suspected DID sufferer's (Bianchi) friends and neighbours, found no early abnormal personality episodes or evidence of abuse, although he himself had engaged in bizarre psychopathic behaviour as an adult.<sup>78</sup> While Bianchi's diagnosis of DID was later retracted, this lack of past evidence is common when such searches have been undertaken. As a consequence, Piper recommends that the disorder be reviewed along with its defining features. He, like others, is not convinced it exists as is commonly understood and perpetuated in the media.

Currently, there appear to be two main schools of thought or 'camps' amongst those who accept DID as a disorder. There are those who accept DID as it is outlined in DSMIV. They appear to classify it as a post-traumatic stress disorder caused by traumatic childhood events.<sup>79</sup> They accept the manifestation of alters and construct their therapy around working with the alters to bring about integration. They deny that this process actually creates or strengthens the different personalities. On the other side, there is a growing group of sceptics who believe DID is a socio-dynamic disorder largely created and re-inforced during therapy. They believe that patient susceptibility and a need for acceptance and attention provide conditions conducive to manifesting alters. They argue there is no evidence to support the presence of alters prior to therapy, and there is little confirmed proof of extreme abuse. Further, police investigations have failed to confirm patients' claims of ritual killings, cannibalism or children used as breeding machines.<sup>80</sup> These appear to be false memories engendered by continuing treatment using hypnosis, suggestion and regression therapy. In addition, Lilienfeld et al found that a high proportion of DID diagnoses came from just a handful of therapists, all of whom were either members of the International Society for the Study of Multiple Personality and

---

<sup>78</sup> See R. Allison's testimony on Bianchi, 'Difficulties diagnosing the multiple personality syndrome in a death penalty case,' *The International Journal of Clinical and Experimental Hypnosis*, Vol. 32, No. 2, 1984, pp.102-117.

<sup>79</sup> David Gleaves, 'The Sociocognitive Model of Dissociative Identity Disorder: A Re-examination of the Evidence' in *Psychological Bulletin* 1996. Vol. 120, No.1, pp. 42-59.

<sup>80</sup> See Lilienfeld et al, Piper and Merskey, Hacking, and Robert Carroll's site on MPD <http://skepdic.com/mpd.html> for reports of bizarre memory recall. Some patients and their families have successfully taken their therapists to court for creating false memories of abuse.

Dissociation (ISSMD) or had expressed an interest in MPD.<sup>81</sup> This would add weight to claims that the condition is an artifice of the therapeutic method.

However, what neither camp denies is that those diagnosed with DID are deeply disturbed, manifest several profound psychoses prior to therapy, are frequently depressed and have suicidal tendencies. Those that eventually seek psychiatric support report unexplained periods of amnesia.<sup>82</sup> The sceptics use these symptoms to claim that DID is not a distinct disorder but a collection of psychoses. On the support side, there is some evidence that comparative testing on patients diagnosed with DID reveal some unique features consistent with background trauma and an attentional deficit. A comprehensive study by Scropo et al demonstrated that there were significant differences in the responses of DID patients to several standard tests for dissociative experience, childhood trauma and symptom inventory compared to those with posttraumatic stress disorder (PTSD) and borderline personality disorder (BPD).<sup>83</sup> They took their findings to indicate that DID can legitimately be viewed as a separate disorder.

On the counter side, both Hacking and Lilienfeld et al cast doubt on the reliability of the testing mechanisms used in these cases. The control groups were college students who also scored highly on the attentional deficit scale. Those others tested had already been diagnosed with DID and at no time was (or has) a random sample been taken of the non-diagnosed population to see how wide the spectrum of responses might be. Hacking also queried the validity of the test mechanisms. While it is acknowledged that those patients diagnosed with DID are suffering from obvious mental health issues, it is by no means clear that DID itself exists as defined by DSMIV or as diagnosed by therapists, especially independently of the therapeutic situation. This casts doubt on the existence of multiples as described in the literature and weakens the impact of DID on demonstrating the natural *disunity* of the self.

---

<sup>81</sup> Lilienfeld et al, 1999.

<sup>82</sup> See Gillig 2009, Gleaves 1996 and Hacking 1995.

<sup>83</sup> Joe Scropo, Sanford Drob, Joel Weinberger and Paula Eagle, 'Identifying Dissociative Identity Disorder: a self report and projective study,' in *Journal of Abnormal Psychology*, 1998, Vol. 107, No 2, pp 272-284.

## 4.9 Implications for the singular self

Whether or not DID can be said to really exist or is taken as a distinct disorder, and whether or not one agrees about its aetiology, there do appear to be people who claim to now live as multiples or to have once been multiples. They claim to experience themselves as either a discontinuous consciousness, with moments of amnesia, or to be one of several distinct personalities sharing the same habitation. In the latter case, they claim not to be deluded but to be genuinely separate personalities. Even if the phenomenon turns out to be iatrogenic, the fact that the self can splinter this way or be multiple appears to work against the claim for the fundamental unity or singularity of the self. It, thus, needs to be addressed within any theory of the self.

At first glance, the existence of DID/MPD appears to represent some of the classic philosophical positions that have led to scepticism about the self. Multiples easily fit the Humean bundle of experiences or collection of mental events models of what is behind selfness; they reflect Parfit's views on personal identity and the absence of self; and the existence of multiples were part of the justification for the Dennettian 'multiple drafts' model of the self, discussed in detail in the next chapter. DID/MPD appears to support the description of the self as a commonwealth with many members, rather than the more commonly experienced singularity. One multiple even described themselves as a 'nation,' denying they are just 'puzzle pieces of some greater whole.'<sup>84</sup> In another sense, the existence of multiples is like a real-life illustration of Locke's hypothetical case of the prince and the pauper, or Parfit's brain-swapping examples.<sup>85</sup> The change from one personality or alter to another is like the moment the transition occurs from the prince's mind in the prince's body to the pauper's mind in the prince's body, or Mr Robinson waking up in Mr Brown's body. Each alter claims they are a distinct and separate identity in the same way the pauper would claim to be distinct from the prince, despite being in the prince's body. Similar to the transition from one multiple to another, the pauper would be amnesiac of the prince's life. The pauper's mind would have no

---

<sup>84</sup> V of the Anachronic Army, *Why We Do Not Have MPD/DID*, <http://www.astraeasweb.net/plural/v.html>

<sup>85</sup> John Locke, Bk II, Chap. Xxvii, s 8 in *An Essay Concerning Human Understanding*, and Parfit, 'The unimportance of identity,' in *The Oxford Book of the Self*, edited by Sean Gallagher, OUP, 2011.

memory of past events in the prince's life and only continuity of memories with his own past life as a pauper. The purpose of Locke's analogy was to demonstrate that one's identity is constituted by psychological continuity or continuity of consciousness, rather than continuity of one's body. DID becomes a dramatic way of demonstrating Locke's point, that personal identity is not intrinsic to the body but to the mind and/or consciousness. It seems to demonstrate that there is no intrinsic self and that personal identity is just the occurrence of events in a brain.

However, on closer examination, the phenomenology described by the multiples appears much closer to everyday phenomenology than first appears. Let's examine the supposed strongest case; several very discrete entities co-existing in the one brain, each of which has fully-fledged self-consciousness and temporal extension during its period of active agency. Each alter expresses awareness of itself as a self-conscious being, has its own personality and name, appears to have past memories and has plans for the future. One could say that the only difference between this account of a self-like entity and our own experience is that we claim to have only one consistent personality at a time. Yet, even this feature appears to be shared by 'multiples.' Multiples are not normally aware of their alters. As far as the currently conscious 'alter' is concerned, they are a singularity, the only 'person' there is in their body. Even after discovering there are 'other personalities' who take over the use of their body, they are rarely (if ever) co-conscious with them. Thus, the self-phenomenology of a multiple may be very close to our own. One could also imagine that if one were to regularly suffer from periods of amnesia where one forgot one's personal history for a period of time (this is what happens to amnesiacs), one might have very different experiences during those episodes of discontinuous identity. If this occurred in conjunction with being in a different environment that demanded different attributes or which enabled different opportunities, it is likely that an alternative personality with different skill-sets could develop. This situation has occurred with those who suffer from Dissociative Fugue.<sup>86</sup> A person with

---

<sup>86</sup> For example, 'The length of a fugue may range from hours to weeks or months, occasionally longer. During the fugue, the person may appear normal and attract no attention; [they] may assume a new name, identity, and domicile and may engage in complex social interactions.' Dissociative Fugue Symptoms, *Psych Central*. Retrieved October 9, 2013, <http://psychcentral.com/disorders/dissociative-fugue-symptoms/>

sudden memory loss can go on to create a whole new life for themselves, even to developing a different personality. Genes aside, what or who we become may be very much a result of environmentally salient features drawing on those capacities and dispositions that fit the situation. With a different background or suddenly challenged by war, one may discover traits in oneself or one may develop skills that would not have surfaced in another environment. Multiples just seem an extreme case of what could happen to any of us, given variant enough circumstances and a ‘clean slate.’

However, as I have argued in the preceding chapters, personal identity is not the same as selfhood. It is the norm for one’s identity to change over time, even if within certain environmental or genetic constraints. One may (and often does) have a sense of being the same self over time. At a minimum, one will have the sense of being a particular singular self. One can still retain a sense of *oneself*, of being a self or a ‘me,’ even if it is not of the same person or identity. One can know that one is (this manifestation, this conscious being), without knowing who one is (Sandra, a lecturer). This concept of the self is compatible with chaotic or even discontinuous identity, such as is seen in cases of schizophrenia, amnesia, Dissociative Fugue, or possibly DID/MPD as discussed here. In these cases, self and identity seem to come apart, whereas in normal circumstances, one’s personal identity is continuous with one’s self over time. Even in DID, each present self still appears to be singular rather than multiple; serial identities rather than multiple selves.

Having said that, if we accept the strongest account of multiples, each alter appears to have a sense of themselves as a conscious ‘I’ while denying sameness of person with the other alters. If true, this could be seen as a genuine case of serial *selves*, rather than serial identities. There is no sharing of memories with other alters, no sharing of thoughts or consciousness, at least insofar as the claimed phenomenology. The ‘speaking alter’ claims to have no sense of continuity with the other alters (although this could also be the case with amnesiacs). There is a distinct subjectivity with apparent temporal extension. Interestingly, Velleman and Strawson both claim that multiple selves are the

norm.<sup>87</sup> They both put forward the idea of timeslice selves, selves that only exist in moments of consciousness, that are discontinuous with each other, hence multiple. There is, however, an important distinction between their account and that of multiples. Multiples, unless integrated, do not appear to be able to engage experientially with their other selves (or past selves). While Strawson argues that his own self\*-phenomenology lacks temporal extension, he can, nevertheless, engage with his past selves\* via memory.<sup>88</sup> Using their terminology, Strawson and Velleman both remember events that their person enacted, even if the action was performed by a past self\*. Multiples, on the other hand, actually claim to be discontinuous or ‘disrupted’ identities or persons who nevertheless are the identical-with-themselves person over time. They claim temporal *extension* but a discontinuous consciousness. When they (the identical-with-themselves person) re-emerge, they claim they are the same self or person they were before they suffered amnesia. This is why they are designated ‘multiples.’ They identify by name, can be called back, remember past actions, and exhibit a reasonably consistent set of characteristics that fits their persona. If all this is true, one would have to classify each of them as a distinct persona with a separate *self*. Both Strawson and Velleman argue for multiple selves because they claim that selves have no intrinsic continuity. With DID/MPD, multiples claim that they *are* extended over time; they have diachronic identity. If we accept what multiples report, their experiences are much more like Cartesian selves than they are momentary selves.

The occurrence of multiple ‘selves’ does not, in itself, argue against the existence of a self. Each self (or alter) is singular and autonomous, has agency and a measure of continuity. For something to count as an object in its own right, it needs the *possibility* of continuity, whether it is actualized or not. Some alters appear to last more than a few minutes and, in some cases, appear to last months or even years. At the time of their manifestation, they have a singular stream of consciousness. Serial selves are not incompatible with the existence of a singular self.

---

<sup>87</sup> See Velleman’s, *Self to self*, 2006 and Strawson, ‘The self and the SESMET’, 1999 for discussions of temporal timeslice selves, referred to in Chapters 3 and 6 respectively.

<sup>88</sup> Strawson introduces the ‘I\*’ and the ‘I’ to distinguish between the temporary ‘I\*’ of the self and the continuous ‘I’ of the person. The I\* lacks temporal extension.

#### **4.10 The case of multiples does not argue against the unity of selves**

In reality, the case of DIDs does not appear as striking as some of the literature would indicate. The evidence from extensive literature reviews like Piper, Lilienfeld et al and Hacking, does indicate that multiples do not exist in the way they are portrayed. It is highly likely they are the iatrogenic product of hypnotic suggestion on highly susceptible minds or a sub-conscious form of attention-seeking behaviour. The case of Bianchi, a serial killer, indicates how psychopathy can be misdiagnosed as DID, especially if the patient is facing execution.<sup>89</sup> He later admitted that his bad alter-ego 'Steve,' who seemingly manifested under hypnosis, was a way of describing his state of mind at the time of the murders. In his case, his 'alters' were aspects of his disturbed personality that came to the fore periodically to commit violent acts. He gave them names to escape culpability. It is likely other DID alters may manifest for similar reasons, especially given a dysfunctional background and underlying psychoses.

In the majority of cases where there appear to be personality disorders that manifest as genuine alters, most are not robust enough to warrant the title of a self or person or distinct personality. They are transient, often momentary, ill-formed and with little autobiography. These momentary 'selves' can run into the thousands.<sup>90</sup> Even in cases where the personalities are slightly more robust, the manifestations are often simplistic stereotypes, what would usually be termed 'alter-egos,' like the inner child, the mother, the sex kitten, the punisher or protector. The personalities have none of the subtle nuances and complexities of an ordinary human psyche. Differences between the resident personalities are also often coarse-grained, such as male/female; left-hand/right-hand; gay/straight; religious/alcoholic. According to Spiegel, DIDs are really a case of having not enough of a self rather than too many.<sup>91</sup> None of the pieces add up to a fully-formed robust multi-dimensional personality.

---

<sup>89</sup> See R. Allison's testimony on Bianchi, 1984.

<sup>90</sup> See Piper and Merskey, 1994 and Hacking, 1995 for a discussion of the spectrum of MPs.

<sup>91</sup> Spiegel was the Chair of the dissociative disorders committee for DSM(IV), Letter to the Executive, *News: International Society for the Study of Multiple Personality and Dissociation*, Vol. 11, 1993.



Whether multiples exist or not, there are some diagnosed DID sufferers who claim to be living as multiples. They do not accept that they are delusional. According to their own self-reports, they really are multiple systems regardless of being caused by trauma, therapy or just naturalistically occurring. It is likely that, despite their denials or affirmations, such people were traumatized, are highly imaginative, and are possibly psychotic or even schizophrenic. If they have always been 'multiple' then they already fall outside of the DSMIV essential criteria of someone with DID. It also means that they are co-conscious as they are aware of their 'alters.' It is likely that their alters are just robust manifestations of different aspects of a complex human personality, similar to Bianchi, where parts have been suppressed or hidden as a consequence of poor parenting. If they are schizophrenic, have other DID's and/or are epilepsy sufferers, then they may have a tendency to dissociate parts of themselves. Symptoms of these conditions can include 'amnesia, fugue, depersonalization, derealisation and identity change.'<sup>92</sup> Hypnosis therapy on damaged and fragile identities, on people who may be prone to delusions, could lead aspects of their personalities which are hidden, dormant or suppressed through trauma, to come to the surface in the form of distinct identities. Hypnosis and regression therapy have been known to cause powerful false memories.<sup>93</sup> The consistent similarities in the kinds of abuse reported under hypnosis are genuine cause for concern in this regard, especially when no supporting evidence can be found. These cases present strong evidence of iatrogenesis of a dangerous kind and cast doubt on the reliability of both patients and therapists in these circumstances. For these reasons, DID's as described in the literature do not have the evidential strength that they first appear to.

The fact that extreme and violent trauma can cause a breakdown of someone's identity or that hypnotic suggestion can produce bizarre effects in an already fractured and damaged human being is not surprising. In and of itself, it does not imply that the self is either non-existent or non-unified in its natural state. Many DID's claim that they were

---

<sup>92</sup> E. Bowman and P. Coons, , 'The differential diagnosis of epilepsy, pseudoseizures, dissociative identity disorder, and dissociative disorder not otherwise specified,' *Bulletin of the Menninger Clinic*, Vol. 64, no. 2, 2000, p.165, pp. 164-180.

<sup>93</sup> <http://skepdic.com/mpd.html>

able to be unified through therapy and that they are now aware of their alters, have dialogues with them and can absorb them into their dominant alter. This would argue against them being primarily separate entities. It lends support to the idea that the basic state of the self is to be unified rather than diffuse. Gillett commented many years ago that commissurotomy patients exhibit an extraordinary level of unity, given the radical separation of the hemispheres.<sup>94</sup> He claims that, rather than revealing disunity, the evidence reveals an overwhelming tendency towards unity as the natural state. It is only in very extreme and unusual circumstances that this unity falls apart. This seems to be the case with DID. If the accounts of abuse and trauma are real, it is surprising that *any* coherent self has survived. The fact that it has indicates how robust the unity and singularity really are.

#### **4.11 Other DIDs and their impact on self-world relations**

I have argued in earlier chapters that selfness or a sense of self can be separated from one's personal identity. I have argued this to demonstrate that what we take as indicative of the self is more than its identity, hence talking about identity will not address all the questions about the existence of the self. I have suggested that one can retain a sense of self, a sense of being a singular conscious subjectivity, distinct from others, without retaining a sense of one's particular identity. Amnesiacs, multiples and schizophrenics could be said to support this position.

If the sense of self is produced by other than one's identity and memory, then one could also imagine the reverse situation; where one lost that sense of self and yet still retained knowledge of one's identity or self narrative. In fact, some very specific types of neurological disorders appear to produce exactly this type of phenomenology. According to Damasio, ailments like anosognosia and asomatognosia (commonly the result of stroke or other brain injury) can damage the sense of self in quite profound ways. Anosognosia and asomatognosia are conditions in which the patient is unaware that they have a physical impairment, such as paralysis, to certain parts of their body. They may ignore or disown the damaged part of their body. Patients suffering from these ailments

---

<sup>94</sup> Grant Gillett, 'Brain Bisection and Personal Identity' in *Mind*, Vol. 95, 1986, pp. 224-229.

tend to demonstrate no *self*-interest yet they retain their full autobiographical history.<sup>95</sup> They can relate personal details and access memories about their holidays or children, but have no sense of an inner self or appear detached from that self. Sufferers of dissociative or derealisation disorders like these can doubt the boundaries of their own bodies. They no longer accurately identify who or what they are as an object in the world. Their sense of themselves as a limited defined entity seems to be lacking. This even extends to their mental world. These ailments indicate that the most fundamental aspect of being a person can come apart – that sense of being a unique individual entity or self.

As with schizophrenia and DID/MPD, these dissociative disorders appear to have implications for the existence of the self. Their manifestation appears to affect the way the patient relates to themselves, their body and their environment. The hallmark of their condition is the presence of what are called dissociative delusions. The delusions defy rationality. While patients may appear to hold rational beliefs in all other areas of their lives, they readily accept bizarre explanations for odd phenomenology, even despite continued contradictory evidence. Patients with Fregoli's syndrome, typified by misidentifying strangers as significant others, accept the validity of their claims despite evidence that they must be mistaken.<sup>96</sup> This is also true of Capgras sufferers. The Capgras sufferer will claim that familiar objects or people in their close environment have been replaced by identical copies.<sup>97</sup> This has led to claims that family members are stealing their possessions to claims that family members themselves have been replaced by look-alike impostors, robots or even aliens. Even though they may acknowledge the bizarre nature of their own statements, this does nothing to change the belief or their behaviour towards the person they claim is an imposter. In contrast, patients suffering

---

<sup>95</sup> Damasio claims that patients with anosognosia retain all their personal biographical details, their 'self narrative' if you will, yet cannot use that information to make decisions about the sorts of things that matter to them. They don't know how they feel or what they want. They are largely indifferent to themselves. Antonio R. Damasio, Chapter 7, *Descartes' Error: Emotion, Reason and the Human Brain*, Papermac, London, 1996.

<sup>96</sup> T. E. Feinberg and J. P. Keenan, 'Where in the brain is the self?' *Consciousness and Cognition*, Vol. 14, 2005, pp. 668-669.

<sup>97</sup> T.E. Feinberg, 'Neuropathologies of the self: Clinical and anatomical features,' in *Consciousness and Cognition*, Vol. 20, 2011, pp. 75-81.

from ailments like prosopagnosia, do not express accompanying false beliefs or delusions.<sup>98</sup> Prosopagnosia is a condition that affects the sufferers' capacity to recognise faces, even though they can still recognise individual features of faces. Unlike in Capgras, the bizarre perceptual experience does not lead patients to make delusional claims. Instead, they acknowledge that there is something wrong with the way they are perceiving the world.

Delusions like those described above are a symptomatic condition of a range of dissociative disorders that represent a class of syndromes seen as neuropathologies of the self.<sup>99</sup> It has been suggested that Fregoli's and Capgras patients suffer from either an over or under 'personal relatedness between the self and the environment.'<sup>100</sup> In Fregoli's, the patient sees familiar others (or places) everywhere. The self/world relation has extended. With Capgras, items have lost their familiarity or personal connection with the patient. It appears that things that were once integrated into their self are now alienated or externalised from that self. Related neuropathologies, like anosognosia, asomatognosia or somatoparaphrenia involve denial of bodily paralysis, non-recognition of one's own body parts or denial of ownership of body parts respectively. These lead to bizarre claims or delusions about their bodies. Cotard sufferers have a related but more general malaise. They claim that they no longer feel part of their environment; they are alienated from everything familiar, including their own mind and body.<sup>101</sup> In all cases, the delusions are monothematic and appear immune to contrary evidence. Rather than acknowledging that they are suffering from an illusion, the patient either denies that anything is amiss or believes the world, including their own body, is as it appears to be to their deluded and distorted senses.

Although uncommon, dissociative disorders are of interest because of the bizarre nature of their delusions and the implications they have for our concepts of self and our belief

---

<sup>98</sup> O. Sacks, *The man who mistook his wife for a hat*, Touchstone, USA, 1998.

<sup>99</sup> See T.E. Feinberg, 'Neuropathologies of the self,' 2011.

<sup>100</sup> T. E. Feinberg, L. A. Eaton, D. M. Roane and J. T. Giacino. p. 383 'Multiple fregoli delusions after traumatic brain injury,' *Cortex*, vol. 35, 1999, pp.373-387.

<sup>101</sup> H Debruyne, M Portzky, F. Van den Eyende, and K Audenaert. 'Cotard's syndrome: a review' *Current Psychiatry Reports*, Vol.11(3), 2009, pp 197-202. <http://www.ncbi.nlm.nih.gov/pubmed/19470281>.

in the veridicality of self-knowledge. Sufferers say things about themselves or the world that are known to be false and can be shown to be false. Yet the patients continue to entertain these bizarre delusions even though they are inconsistent with other beliefs they hold. They remain convinced that the false beliefs are true, despite good reasons to doubt them. Cotard syndrome is of particular interest in this context. Their disturbance seems to be directly related to their perceptions of themselves from the *inside*. They manifest bizarre delusions about themselves, their bodies and their own mind to the point that they claim to not exist. This seems to contradict the standard conclusion we would normally draw from our self phenomenology. The Cogito argues that the act of thinking itself is proof of existence. If Cotard sufferers come to an alternate conclusion, it could be seen to cast doubt on the reliability of the conclusions we draw from the evidence of our senses. Could we all be mistaken and there really is no-one home?

#### **4.12 Cotard's and related derealisation syndromes: their neurological basis**

Cotard's syndrome was first identified by Jules Cotard in 1880, when he discussed the case of a woman who denied the existence of several parts of her body and believed she didn't need to eat. Cotard was investigating an extreme form of melancholia or depression. He distinguished three categories of patient – those with negation delirium simpliciter, those with negation delirium leading to general paralysis, and those with negation delirium associated with persecution delirium (complex alienation). He defined this condition as a 'psychic state proper to the chronically anxious' and identified it as a syndrome.<sup>102</sup> According to Cotard, all patients with the syndrome denied the existence of certain things that they have had at some point (names, parents, children, their age) but some go even further to deny the existence of parts of their body as well (no head, no stomach, or no body). Some of these cases were accompanied by hallucinatory feelings such as their brain turning to liquid, their body expanding, their legs vanishing, no blood or their body becoming something else. In a more recent case, a young man who

---

<sup>102</sup> Daniel Heller-Roazen, *The Inner Touch: archaeology of a sensation*, Zone Books, New York, 2007, p. 267. See also 'Cotard's syndrome: a review' 2009, pp 197-202.

developed Cotard's after a brain injury thought he had died and was in hell.<sup>103</sup> A Filipino woman with Cotard's complained that 'she was dead, smelled like rotting flesh, and wanted to be taken to a morgue so that she could be with dead people.'<sup>104</sup> Sometimes the sense of delusion is accompanied by feelings of distrust or persecution. The most common belief expressed by Cotard sufferers is the delusion they are dead, now known as Cotard delusion. One man with Cotard's tried to commit suicide to prove to those around him that he really was dead.<sup>105</sup> One can see the enormity of the delusions in the following excerpt from Cotard.

In all the patients, the hypochondriachal delirium introduces great delusions: their brain, stomach, heart, blood, spirit and/or body are missing. They are damned, the organs do not exist, the body is reduced to a mere machine. The delusions may include religious, metaphysical and abstract ideas of persecution. To such ideas, delusions of immortality may come to be included. Along with or following the ideas of immortality may come ideas of body expansion in space: they are immense, their dimension is gigantic, they can touch the stars, they may feel possessed by powerful demons, their head expands until it occupies an entire church. At times the body no longer has limits, it extends to the infinite and it disperses in the universe.<sup>106</sup>

One can see from these symptoms that the normal bodily interoception has gone awry. What sufferers lose is their sense of parameter, whatever makes them experience their own body as bounded or solid or as related to themselves in the normal way, what Carruthers referred to as the boundary<sub>b</sub> and boundary<sub>m</sub> processes. This appears to create an abnormal sense of the body and its parts such that they feel transformed or absent. Cotard patients lose touch with their brains and minds as well as their body. There is a disconnection between themselves and their internal experience of themselves. They often refer to themselves in the third person but claim they, the I, does not exist. Yet they continue to use the first-person indexical while claiming not to exist anymore. This depersonalisation or derealisation is not the same as that reported by schizophrenic sufferers who disown their thoughts. The schizophrenic sufferers hand ownership or

---

103 See G. Heffner, 'The Cotard Syndrome Website,' <https://sites.google.com/site/autismhome/Home/miscellaneous/cotard-syndrome>, accessed June, 2011.

104 A. Ruminjo and B. Mekinulov, 'A case report of Cotard's Syndrome,' in *Psychiatry* (Edgmont), Vol. 5(6), 2008, pp 28–29.

105 These stories are well-documented because of their bizarre nature and can be found in many places, but see <https://sites.google.com/site/autismhome/Home/miscellaneous/cotard-syndrome>.

106 Translated from Jules Cotard by G.J. Heffner in 'The Cotard Syndrome website', *Medline*, 1999.

agency of (some of) those thoughts to someone else. It is the thoughts themselves that appear alien. With Cotard patients, it is not specific thoughts that feel alien to them; there is a complete lack of association with themselves, including their own minds such that they believe they must no longer exist. They have a much more global loss of affect which creates a far more alienating experience.<sup>107</sup> They believe what their senses seem to be telling them, that they must be dead or are no longer embodied.

There is strong evidence that these neuropathologies have a neurobiological underpinning. According to Stone and Young, both Capgras and Cotard have a similar aetiology.<sup>108</sup> Both delusions have been known to manifest after a patient has suffered a stroke or damage to the occipito-temporal or temporo-parietal regions of the right hemisphere of the brain. This may explain why patients may alternate between the two syndromes, depending on whether they are depressed (Cotard's) or paranoid (Capgras). The former condition directs the dissociation towards themselves while the latter directs it towards others. Duggal et al claimed that a patient with bi-polar who developed Cotard symptoms showed evidence of frontal and parietal lobe dysfunction, often responsible for a denial of body parts in the patient.<sup>109</sup>

Feinberg also claims that dissociative misidentification syndromes and derealisation syndromes are usually the result of right frontal brain lesions or focal brain pathology (right fronto-temporal injury or right frontal infarct).<sup>110</sup> There is some evidence that Capgras sufferers may also have a history of schizophrenia, although evidence of Cotard's was not as frequent.<sup>111</sup> Some epileptics can have either syndrome and this may be related to the site of the seizures (i.e. temporal lobe epilepsy).<sup>112</sup> The delusions

---

<sup>107</sup> Philip Gerrans argues that Cotard delusion is a reasoning deficit and that depression can cause widespread affective suppression. 'Redefining the Explanation of Cotard's Delusion', *Mind and Language*, Vol. 15, no. 1, 2000, pp. 111-122.

<sup>108</sup> Tony Stone and Andrew Young, 'Delusions and Brain Injury: The Philosophy and Psychology of Belief' *Mind and Language*, Vol. 12, nos 3-4, Sept-Dec, 1997.

<sup>109</sup> H. Duggal, K. Jagadheesan and S Nizamie, 'Biological basis and staging of Cotard's syndrome,' *European Psychiatry*, Vol. 17, 2002, pp 108-109.

<sup>110</sup> Feinberg and Keenan, 2005, pp. 666-668.

<sup>111</sup> See M. Coltheart, R. Langdon and R. McKay, 'Schizophrenia and Monothematic Delusions,' *Schizophrenia Bulletin* Vol. 33 no. 3, 2007, pp. 642-647.

<sup>112</sup> See Feinberg, 2011; and Bowman and Coons, 2000.

themselves are generally ‘monothematic and often circumscribed,’ the nature of the delusion is heavily constrained with little indication of more generalised bizarre beliefs.<sup>113</sup> While there are cases of Cotard sufferers who refuse to eat and neglect themselves, in line with their belief they are dead, most continue to engage in everyday activities, albeit minimally. There is some evidence that continued presentation of counter evidence to someone with Capgras can hold the delusion at bay.<sup>114</sup> However, this change only has a short duration and the delusion returns once the presentation of such evidence becomes less frequent. Patients with Capgras do show paranoid tendencies while Cotard’s sufferers are chronically depressed. There has been some success in treating patients with appropriate anti-depressant and anti-psychotic medication. Electric shock therapy has also been effective in some cases.<sup>115</sup>

Anosognosia causes patients to deny that there is anything wrong with their body even when, for example, their arm is obviously paralysed. As a consequence, they confabulate by making up stories to explain why they didn’t move their arm as requested. Asomatognosia is a non-recognition of one’s body parts as one’s own, often called hemispheric neglect because one side of the body/world is ignored. Somatoparaphrenia is an extreme case of bodily neglect that involves delusional claims about the neglected body part, usually the left arm. For some reason, patients suffer bizarre delusions about this now lifeless left arm (sometimes their left leg, but this is not as common). The patient disowns it so completely that they claim it is a thing (like a rock), someone else’s arm, or even a separate person (like a bad child). They use elaborate metaphors to describe what has happened or is happening to their limb but they never talk about the arm directly. This is how one patient described their paralysed arm, when asked if she had other names for it.

Patient: Her. She belongs to me so she’s a her. She’s mine but I don’t like her very well. She let me down.  
Doctor: In what way?

---

<sup>113</sup>Stone and Young, 1997, p.331.

<sup>114</sup> Coltheart et al, ‘Schizophrenia and monothematic delusions,’ 2007.

<sup>115</sup> A. Ruminjo and B. Mekinulov, 2008.



Patient: Plop plop rock rock nothing. I was on my way home out the door and then she went and did this [pointing to her left arm]. She didn't ask if she could [shaking her head back and forth]. I have to be the boss not her. [she said pointing to her left arm.]<sup>116</sup>

According to the way the patient describes it, the limb takes on an agency and character of its own, independent of the will and control of the owner. Some patients have their limbs removed so they can 'feel whole' again.

It should be noted that the neuropathologies of the self are symptomatically diverse and aetiologically heterogeneous. Coltheart et al claim that there are at least 15 different neuropsychiatric or neuropathological conditions that have generated Capgras delusions.<sup>117</sup> Similar causal variation exists with Cotard's. Some of these conditions are schizophrenia, epilepsy and Alzheimer's. Some sufferers have none of these conditions. This means that the basic aetiology cannot explain the variety of delusions that result from the malfunctions. Nor can it explain the extreme phenomenological experiences or the resulting robustness of the delusions. Patients don't just experience bizarre things; they believe the bizarre conclusions they draw from those experiences. As is often remarked, other conditions with similar aetiologies do not result in similar bizarre beliefs, even if the patient suffers from illusions or hallucinations. Other patients with left side paralysis and right temporoparietal damage do not disown their arm but claim, rightly, that they are paralysed.<sup>118</sup>

In line with Coltheart et al, other neurologists argue that, while the aetiology is diverse, there is evidence of cognitive homogeneity. Feinberg and Keenan claim that a cluster of dissociative disorders (of which Capgras, asomatognosia and Fregoli's are but a few) are commonly the result of right frontal brain lesions or focal brain pathology (right fronto-temporal injury or right frontal infarct).<sup>119</sup> They claim that asomatognosia, for instance, is caused by 'right hemisphere lesion, [leading to] left hemiplegia, severe sensory loss

---

<sup>116</sup> Feinberg and Keenan, 2005, p.667.

<sup>117</sup> Coltheart et al, 2007, p. 645

<sup>118</sup> Coltheart et al, 2007, p. 644.

<sup>119</sup> Such as Delusional Misidentification Syndrome (DMS) and Delusional Reduplication Syndrome (DRS), Feinberg and Keenan, 2005, pp. 666-668.

on the left side, and left hemispatial neglect.’ Similarly, Duggal et al claim that parietal lobe dysfunction leading to a denial of body parts can occur in ‘lesions of the dorsolateral frontal lobe, cingulate gyrus, thalamus and neocortex.’<sup>120</sup> This appears to indicate a neurological causal link between Capgras, Cotard’s and other misidentification or derealisation syndromes (DMS/DRS) that produce bizarre delusions about the body or body parts. Coltheart et al claim that there is likely to be a 2-factor causal pathway for delusional beliefs - one specific to the ailment and the other damage to some part of the right frontal lobe. Frontal lobe damage appears to impact on the ability of the patient to correctly identify and relate to *self*-information.

#### **4.13 Implications for the self, self-knowledge and use of the indexical**

The problem of the self exists because we claim our existence as a self is irrefutable. I think and experience things from my perspective as a conscious entity. If someone with Cotard’s, for example, can still consciously act in the world and yet claim there is no ‘I’ or that they don’t exist, then this appears to directly contradict our claims for direct self-knowledge and our belief in a self, the evidence for which is drawn from our self-phenomenology. Heller-Roazen claims that ailments like Cotard’s cast doubt on the existence of the self as a discrete entity for this reason.<sup>121</sup>

Heller-Roazen compares Cotard’s with phantom limb syndrome, and claims that Cotard’s and phantom limb are the mirror image of each other in the way they manifest.<sup>122</sup> The phantom limb patient claims awareness of bodily parts that no longer exist; the Cotard sufferer claims no awareness of bodily parts that still exist. Both are mistaken. According to Heller-Roazen, they are all experiencing a disturbance of

---

<sup>120</sup> Duggal et al, 2002, p. 108.

<sup>121</sup> Heller-Roazen, *The Inner Touch*, 2007.

<sup>122</sup> There may appear to be similarities in the way the two syndromes manifest but the causal pathways of each are vastly different. Also, Cotard’s is a delusional syndrome whereas phantom limb is an illusion. There is no false belief or denial of the missing limb in the patient, unlike anosognosia or asomatognosia.

‘coeanaesthesia’ or the vital sense.<sup>123</sup> This inner touch is supposed to inform the person (or animal) of its own body, its extension and boundedness, and all that is contained inside. Cotard and phantom limb delusions cast doubt on the reliability of that mechanism or bodily apperception to accurately represent the body to itself. In Cotard’s, the sufferers disown the thoughts and feelings of the body they appear to not be inhabiting. They deny the existence of themselves, as a self, while still acknowledging the existence of some body or other. With phantom limb, the patient’s perception contains an impression of something that does not exist. Heller-Roazen argues that this means the inner touch, the general perception of the extended body, is not a direct perception; it does not inform the brain directly of what is really there and present. This casts doubt on our supposed experience of ourselves as a unified agent. Additionally, patients with depersonalisation syndromes can report feeling alienated from their self such that it seems like a new self with new experiences is taking over. Heller-Roazen claims that if the person can be alienated or separated from the subject who is speaking, then personhood is not necessary for conscious life. Following this line, one could even suggest that if the speaking self is different from the experiencing self, then one or the other is not necessary for consciousness. In fact, persons and selves could be distinct from consciousness. Human beings could live and reason in the absence of their own person or self. Phenomenological evidence from these odd disorders would suggest there could be thinking outside of a self.

I think the above reasoning demonstrates that we need to be cautious what conclusions we draw from some phenomenological reports, particularly when the persons involved are known to be delusional. While Cotard’s and other derealisation or depersonalisation ailments do raise issues about the robustness of the sense of self and the mechanisms underlying that sense of self, I am not convinced that they lead to the sceptical conclusions suggested by Heller-Roazen. Janet recommended that we be cautious about taking extreme phenomenological claims at face value.<sup>124</sup> For instance, Cotard’s

---

<sup>123</sup>Heller-Roazen, 2007, p.247. According to Schiff, coeanaesthesia is ‘the complex of all sensations, which, in any given moment, are more or less distinctly transmitted to our consciousness, and which constitute the content of our momentary consciousness.’

<sup>124</sup> Janet cited in Heller-Roazen, 2007, p.282.

sufferers are not dead; they do not act as though they are dead even though they may believe they are dead (although some Capgras patients have been known to act on their false beliefs). While somatoparaphrenia patients claim to believe that their paralysed arm is really their dead brother or their doctor's arm, this is not the case. Some will alter their claim temporarily when shown evidence to the contrary.<sup>125</sup> One of the difficulties of drawing philosophical claims from phenomenological reports is that we are reliant on the personal accounts of the patients, and their capacity to lucidly and accurately describe their condition. It is obvious that something phenomenologically different is happening with Cotard's patients, for example. Their normal way of perceiving or experiencing themselves is disrupted in such a way as to create this sense of unreality in regards to both their sense of self and their own bodies. If we accept their phenomenological claims as describing what it feels like *for them*, they support the existence of a singular subject rather than the opposite. The patients claim they are not aware of themselves even though they are conscious. Their bodies are rotting or liquefying. They variously claim that 'I think but I don't exist' or 'I am but I don't think.' This would seem contradictory. In actual fact, the conscious 'I' is present and aware of the disintegration of the bodily self. They have not lost their sense of singularity or of existing. Instead, they exist as immortal or in the hereafter (possibly influenced by their theological beliefs). The 'inner' conscious self has become disembodied, detached from both the body and the brain. It exists almost as a pure Cartesian ego. This could be called selfhood in essence, distinct from its personal identity, its history and its social networks, which seem to remain with the body. One could argue that Cotard's actually proves the existence of a mental self, rather than the reverse.

Patients suffering from depersonalisation syndromes sense 'with clarity' that they no longer sense and no longer sense themselves. They express 'feelings of emptiness.'<sup>126</sup> So there is still a strong sense of feeling, of sensation. It is just that the senses (or some

---

<sup>125</sup> Fotopoulou, A., Jenkinson, P.M., Tsakiris, M., Haggard, P., Rudd, A. & Kopelman, M.D, 'Mirror-view reverses somatoparaphrenia: Dissociation between first- and third-person perspectives on body ownership,' *Neuropsychologia*, 49, 2011, pp. 3946-3955.

<sup>126</sup> Heller-Roazen, 2007.

sense) are not sensing the right kinds of sensations or the signals are being misinterpreted or are just missing. Gerrans argues that a breakdown in the sense of affect in patients with delusional misidentification or derealisation syndromes could lead to a loss of sensation as described in Cotard's.<sup>127</sup> Depression seems to lead to a flattening of affect, resulting in less acute stimulation from the environment and a subsequent loss of attachment or engagement. Gerrans claims that affect programs play an important role in individuating a particular organism from others and in individuating that organism to itself. It does this by identifying perceptual information being received by a particular entity as being received by *that* particular entity. This process attributes relevance or meaning to that information. Without such relevance, the entity would not know how to respond or act appropriately to the information received.<sup>128</sup> As Gerrans says, 'in the absence of affective processing, perception and cognition have no bodily consequences and thus are not 'felt' at the phenomenal level to belong to the agent.'<sup>129</sup> If this pathway were damaged, one could imagine that information received through interoception, for example, might lack personal significance. While this doesn't explain the bizarre nature of the delusions, it may help to explain why, in Cotard's for example, their own bodies have become alien to them. This makes it rational, at one level, to claim either that they exist as an immaterial self because they are no longer 'embodied,' or that they are dead.<sup>130</sup>

In line with Gerrans, Stone and Young argue that recognition of familiar objects *as familiar* requires visual information from both the ventral and the dorsal neural route.<sup>131</sup> The ventral route is for overt recognition of objective facts while the dorsal-limbic route provides affective information leading to an emotional response and is 'implicated in the

---

<sup>127</sup> Gerrans, 'Refining the Explanation of Cotard's Delusion,' *Mind and Language*, Vol. 15, no. 1, 2000.

<sup>128</sup> Philip Gerrans, 'Cognitive Disintegration and Delusional Misidentification,' *Psychology*, 1999.

<sup>129</sup> Gerrans, 2000 p. 119.

<sup>130</sup> Gerrans refers to this as an 'observationally adequate belief', 'Refining the Explanation of Cotard's Delusion,' 2000, p. 115.

<sup>131</sup> Stone and Young, 1997, p.337

process whereby “relevance” is attached to an attended object’.<sup>132</sup> According to Feinberg, right frontal lobe damage appears to create problems with identifying things that have relevance for and to the self, such as family, places or parts of one’s body. This can lead to ‘inappropriate alienation’ of self from things or an inappropriate projection of self onto others, including places. This would imply that the right frontal (and temporal) lobe plays an important role in establishing the correct relationships between self and one’s environment or world. Feinberg cites detailed research evidence that indicates that self-recognition is predominantly a right hemisphere function.<sup>133</sup> Experiments he conducted with split brain patients and recognition of self images showed 1.2-1.8 times more activity in the right frontal region than the left (right frontal and right cingulate). Feinberg examined the cortical correlates of self-recognition using fMRI and found that response to self-recognition was still higher than responses to other familiar faces. In addition, the right hemisphere and the medial prefrontal cortex appear to be crucial for monitoring self and other activity. If this is the case then it would explain why damage to the right frontal cortices may lead to unusual symptoms in relation to self/world or self/self recognition. This could explain the bizarre descriptions used by somatoparaphrenia patients to describe their paralysed arm. Given the lack of lingual capacity in the right hemisphere, the left side of the brain may well have to resort to odd kinds of metaphorical imagery to try to communicate the loss, given its supposed lack of awareness of the loss.<sup>134</sup> Similar confabulation can be found in split-brain patients who also use elaborate techniques to communicate across the hemispheres.<sup>135</sup>

There is strong evidence that dissociative or depersonalisation disorders have a neurological trigger, usually a stroke of some kind that damages certain sensory pathways or particular (right) lobes of the brain. This is likely to be multi-factorial as suggested by Coltheart et al and Feinberg and Keenan. Patients develop their various

---

<sup>132</sup> Young refers here to Bauer’s model of visual processing. Bauer claims the dorsal pathway has ‘multiple functions encompassing automatic emotional responses to stimuli which have personal relevance.’ Andrew Young, Chap 10, *Face and Mind*, OUP, Oxford, 1998 p.265.

<sup>133</sup> Feinberg, 2005, pp 673-4.

<sup>134</sup> Dennett refers to this phenomenon as an epistemic absence, one is not aware that the information is missing so one can’t recognize a loss. Dennett 1991.

<sup>135</sup> R. Sperry, M. Gazzaniga and J. Bogen, ‘Interhemispheric relationship: the neocortical commissures; syndromes of hemisphere disconnection,’ *Handbook of Clinical Neurology*, Vol. 4, 1969.

conditions over time or in conjunction with other pathologies. They all express an awareness of phenomenological difference, a sense of missing something or erroneous sensations that are new that lead to delusional beliefs. They know what they are missing or what has changed because it was there before. With Cotard's, they miss the experience of themselves as embodied subjects. This is interesting because exteroception and proprioception, the outer and inner perception of the body that utilises such things as skin sensors, muscle tension, tendon stretches, blood pressure, are still intact, because general motility and coordination are unaffected. A person who has a disorder affecting proprioception will be unable to walk or even hold a cup.<sup>136</sup> Damasio claims that patients with anosognosia lack the capacity to make informed or rational decisions about things that relate to their lives. They lack *self*-interest. Yet, they have no loss of memory or loss of autobiographic details. Their sense of self is impaired without loss of personal identity. The same is true of Cotard sufferers. It is the 'I' that detaches from its physical realm, whereas knowledge of one's identity remains but is distinct from the perceiving subject. This would indicate that the sense of self comes from a deeper, more primitive apperception of the self. This apperception appears to produce the sense of embodiment or inner sense referred to by Heller-Roazen. However, the sense of being an 'I' remains unaffected. Consciousness of oneself must be produced by an even more primitive mechanism. This makes it likely that one's sense of self is driven by some basic neurological mechanisms that process self-information and produce our self-phenomenology, underpinned by consciousness.

In Chapters 2 and 3, I referred to ipseity or our sense of being. What Heller-Roazen calls the inner touch or common sense is a less specific, more generalized sense of embodiment.<sup>137</sup> This is more than the constant daily perceptions of objects or specific internal states. While these are self-identifying and self-informative, our direct awareness of them ceases in certain circumstances, i.e. during sleep or when engrossed in activity. What both Heller-Roazen and Damasio refer to is a bodily-focused background hum or sensation that is always present, albeit not in the forefront of

---

<sup>136</sup> See Sacks for examples of proprioceptive disorders and their consequences, in *The Man Who Mistook His Wife for a Hat and other clinical tales*, Summit Books, USA, 1985.

<sup>137</sup> Daniel Heller-Roazen, 2007.

consciousness.<sup>138</sup> The writers that Heller-Roazen refers to appear to be striving to articulate this sense; rather like a sense of embodiment *from the inside*. It is a form of perception, but it is not like the perception of our sense organs because there is no specific organ that does the perceiving or a specific site at which the perception exists. It is thus general and unlocalised, like the sense of touch. It could be described as the beat of the living organism. Neurological accounts like Damasio's and Feinberg's are showing that there may be some reality to this phenomenology. According to Damasio, there is a range of mechanisms that are responsible for producing a phenomenological sense of being. The limbic system, for instance, produces a constant, continually updated representation of the body's states that provides a kind of background sensation to all conscious states.<sup>139</sup> This not only gives a sense to the subject that these states are its own but it also provides the sense of subjectivity that one is *in* a particular state. It is like a continual state of engagement with one's internal world. This condition is independent of one's personal identity or self-narrative, although it may well provide the content to one's self-concept. It is primitive and most likely prior to the development of one's personal identity. Damasio calls this the proto-self, the first stage of the self. This is one's sense of the living body, of being a living entity. It provides the sense of constancy and continuity that binds the Parfitian mental events and personal experiences together as belonging to a single entity, what Damasio refers to as the 'core self.' It is this core self which is disturbed and damaged by DMS/DRS. Even here, however, the subject remains intact, if deplete of many of its defining features. It becomes truly a basic self.

#### **4.14 The self in the brain**

Neuropathologies of the self do raise questions about what our phenomenology is showing us and how reliable it is. This does not mean that the self does not exist and our sense of self is an illusion. My dead arm is not an illusion, even if I can no longer sense it or claim it as my own. Neither does the fact that I can stop owning a part of my body necessarily uncover an underlying disunity of self. What it does reveal is that the experience of unity and the phenomenological sense of self are vulnerable. They rely on

---

<sup>138</sup> Heller-Roazen, 2007; Damasio, 1996.

<sup>139</sup> Damasio, 2010, p. 107 and p.180. I discuss his account in detail in Chapter 7.



the correct functioning of underlying neural mechanisms. If those mechanisms are damaged too much or malfunction, then the sense of self starts to disintegrate. What this means is the self can be damaged or destroyed like any other object. While this doesn't in itself prove that the self is a real unified entity, it doesn't show that it is not. Take the legs off a table and it will no longer function as a table. It appears that the correct functioning of certain neural mechanisms is to individuate and unify one's self, thereby producing a consistent sense of self. Ailments like Cotard's and anosognosia demonstrate how fundamental our sense of embodiment is to the normal functioning of the self. Without it, the self does not function correctly and has no reason to act; perceptual experiences have no meaning. This also indicates that the self plays a functional role within the human cognitive system in ways that our identity or personality do not.

I want to demonstrate that the self can be considered a real existent thing by providing evidence of underlying physical mechanisms that appear to be intimately connected to many of our self-like faculties and their corresponding phenomenology. Consequently, I argue that there is a physical correlation between the sense of self and those physical structures. This can be demonstrated by the loss of self-phenomenology when those mechanisms are damaged or destroyed. The concerted actions of those mechanisms produces our self phenomenology or sense of being something (a self) and could qualify for being a self, if there were enough unity and cohesion in the way they interacted with each other. One does have to be careful about claiming that certain neural networks (N) are responsible for producing certain capacities (C), just because damage to N is present when C is lost or compromised.

This may not mean that N causes C or that N is responsible for producing C. C may be a side effect of something else and the damage to N may be symptomatic of something else. Likewise, C may only occur when other causal or contributing factors are present as well. Claims about dysnarrativia are a case in point. Some narrativists argue that patients with dysnarrativia, the inability to comprehend and use speech due to damage to

the language faculty, lose their self, as I discussed in Chapter 3.<sup>140</sup> They use this finding to indicate the role of the narrative in self-construction and its loss as evidence of self-destruction. However, as Zahavi points out, patients with advanced Alzheimer's or Korsakoff syndrome suffer a range of debilitating cognitive complaints in addition to those affecting story-telling.<sup>141</sup> Even if these patients can be said to have no self, one cannot claim that this was due to the loss of narrativity.

Papineau claims that the weak correspondence between mental level disorders and physical disorders has led to the commonly-held conclusion that some mental disorders are just 'mental' disorders and, therefore, do not constitute real illness (there is no obvious underlying physical disorder). One can have a perfectly healthy functioning brain (and some patients do) and yet be depressed.<sup>142</sup> This raises the issue of mental/physical correlation or reduction. Even if a particular physical pattern *p* occurs in (almost) every case of *x*, this does not mean that *p* causes *x*. Even if *p* occurs in every case of *x*, and it may be there is no instance of *x* without *p*, it could still be the case that some other factor *f* causes both *p* and *x*, or that *f* needs to be present in addition to *p*. Coltheart et al claim just this with monothematic delusions such as Capgras.<sup>143</sup> Capgras patients may be delusional because of several coinciding neurological factors, none of which explain why they limit their delusion to certain significant people, nor why they are paranoid. This means that there may not be a 1:1 correspondence between specific physical structures and the production of certain phenomenology. I return to this point in more detail in the next chapter, Chapter 5.

What I can claim from the neurological evidence is that the unusual phenomenology is likely to be produced because there is damage to mechanisms in the brain that are known to 'take care of' some very specific functions relating to self-identification and discrimination. There may be no choice but to experience *x*, *y* or *z*, hence the consistency of the delusions. In a parallel vein, there will be little choice about the kind

---

<sup>140</sup> Kay Young and Jeffrey Saver, 'The Neurology of Narrative,' *Substance* vol. 94/95, 2001, pp 72-84.

<sup>141</sup> Zahavi, 2007, p. 185.

<sup>142</sup> Papineau, 1994, pp73-84.

<sup>143</sup> Coltheart et al, 2007.

of self-phenomenology we develop as a result of the normal functioning of certain parts of our brain. Our sense of self will be produced by the actions of self-related mechanisms. This argues against a purely socio-cultural explanation for the nature of the experience. However, I do not want to claim that the self can be reduced to the presence of certain structures or processes in the brain, in the way that Carruthers does or Parfit, such that this negates its existence. Even though the correct functioning of these mechanisms is responsible for producing much of our self-phenomenology, they appear to work in concert with each other. Some self-like features or capacities appear to emerge from that integrated function. One could claim that this is how the self is constituted; it is what the self is.

In the next chapter, I outline two models of the self that put forward complex neurological explanations for the existence of the self, much as I am advocating here. However, both Dennett and Metzinger argue against the reality of the self and our self-phenomenology. I will present their positions in detail and demonstrate why I think they don't work as an explanation of the self. I argue that denying phenomenal consciousness leads to real world scepticism. I want to claim that there are better models of the self that are still physicalist but which do not require us to eliminate our self-phenomenology.



## CHAPTER 5

# PHYSICALIST CONCEPTIONS OF THE SELF: DENNETT'S MEME MACHINE AND METZINGER'S PHENOMENAL SELF MODEL

No science can state what the Self is without stating it in perfectly general terms... (what the self is not) - truth lies in the actuality of existence as an individual subject, not in the objective generalities of science. 1

### 5.1 Physicalist conceptions of the self and phenomenology

In the last chapter, I argued that the pathologies of the self do not necessarily argue against the existence of a self, nor against a unified or singular self. Instead, what the evidence from pathologies appears to demonstrate is that there are a series of very specific areas of the brain that are responsible for producing many of our self-capacities. When these are damaged, there appears to be a direct impact on our self phenomenology and sense of self. This would support the idea that the self is a manifestation of the correct functioning of a self-related neural network that is responsible for creating our sense of self and its phenomenology.

In this chapter, I want to discuss two physicalist accounts of cognition that describe how the brain creates that mental world and all its subsequent phenomenology. However, unlike the view I put forward in the final chapter, both accounts claim that what is produced is actually an illusion. While they argue that the mind, self and consciousness are *products* of the brain, they claim that this fact itself demonstrates that none of these things are fundamental features of our cognitive architecture. At its core, the brain is a non-conscious mechanism. It creates, either indirectly or through a change in its architecture, the phenomenology that leads us to posit the existence of a mind and a self-conscious self. The mind and self, with all its phenomenology, are just epiphenomenal illusions or hallucinations.

In the following sections, I discuss an account of the self presented by Dennett in which he explicated the notion of the self as the Centre of Narrative Gravity.<sup>2</sup> I

---

<sup>1</sup> Søren Kierkegaard, , *The Concept of Anxiety* trans. Reidar Thomte, Princeton, NJ, Princeton University Press, 1980, p.78.

<sup>2</sup> Daniel Dennett, *Consciousness explained*, Little, Brown, Boston, 1991.

follow this with a more recent version presented by Metzinger.<sup>3</sup> I use Metzinger because he appears to extend Dennett's framework to justify his own phenomenal self-model (PSM), while representing a more detailed computational information-processing model of the mind and self without recourse to Dennett's meme machine. I have chosen Dennett and Metzinger because they are a good illustration of a particular body of research within philosophy of mind which draws on current computational and neuroscientific research. Both accounts posit their version of the self without adverting to a Cartesian self. They both proffer explanations for the existence of the self within a physicalist framework. Like other writers within this field, they claim that self-consciousness is a late evolutionary development, superimposed on a non-conscious framework. This leads them to a sceptical approach to both the existence of a self and its phenomenology. That scepticism rests on a dismissal of phenomenal consciousness as it is experienced and, as a consequence, a dismissal of our phenomenological reports about what it is like to be a self-conscious entity.

I argue that neither Dennett nor Metzinger adequately address the phenomenology of the self nor do they provide an explanation for its presence. The cognitive system they posit could operate equally well whether phenomenally conscious or not. I also argue that they do not successfully demonstrate that the phenomenology is illusory or that we are not really conscious systems. In fact, too much scepticism towards the reliability of phenomenological reporting can lead to epistemic scepticism about all things, not just the self. I go on to claim that there is neurological evidence that the phenomenology itself is not just an inference or judgement but is an actual product of certain specific neurological mechanisms in the body and brain such as the limbic system or parts of the hippocampus, as intimated in Chapter 4. This makes phenomenal consciousness and selfhood more likely to be naturally-occurring features of our type of functioning cognitive architecture. It also appears that self-consciousness may exist at a much more fundamental level than either Dennett or Metzinger intimate. I conclude by claiming that there are better explanations of the self and its phenomenology than those presented by Dennett and Metzinger that would still qualify as physicalist. I end with a brief discussion of the implications of physicalism on the ontology of selves and other mental states.

---

<sup>3</sup> Thomas Metzinger, *Being No-one: The self model theory of subjectivity*, MIT Press, Mass., 2003.

## 5.2 Dennett's explanation for the existence of the mind, self and consciousness

I want to make it clear at the start that Dennett does not claim that there are no such things as 'selves'. He is not strictly an eliminativist like Churchland, although he has been accused of this.<sup>4</sup> I address this issue in a later section of this chapter. According to Dennett, "'selves' exist and each person, more often than not, has (at least) one."<sup>5</sup> However, according to Dennett, these selves are not like the metaphysical self espoused by Descartes. The kind of self he posits is not a substance or an independently existing entity. It is a fabrication, a story about a self. The term 'self' is used as an epithet to describe the behaviour of a system that not only appears to others as if it has a self but which also (generally) believes itself to have a self. Each human self is a construct produced over time by the changes wrought in the existing cognitive architecture as a result of exposure to complex social information. According to Dennett's ontology, selves do exist as real discernable patterns of behaviour and, as a consequence, can be considered 'real abstractions.'<sup>6</sup> However, they have no ontological reality. They are not real objects. There is nothing more to the self than the story we tell about that self as an explanandum for our behaviour. The self is the fictional story-teller and protagonist of the self-narrative. The phenomenology of the self, or the sense of self, is created by that narrative. This is the effect of having certain 'self memes' or beliefs about our self, available in our particular culture.<sup>7</sup>

Dennett's substantive thesis is a theory of consciousness. However, one cannot address consciousness without also addressing the supposed existence of the self, as

---

<sup>4</sup> Paul Churchland controversially claimed that talk of qualia such as pains, tastes, colours could be eliminated because they could be more accurately described by referring to underlying causal mechanisms such as 'C fibres firing,' and that a mature neuroscience would demonstrate that there was no 1:1 intertheoretic reduction. See *Matter and consciousness* (revised edition), MIT Press, Cambridge, 1988, pp. 36-49. Dennett holds similar ideas about propositional attitudes, expressed in *The Intentional Stance*, MIT Press, Cambridge, Mass., 1987. I discuss this point in more detail in the final section of this chapter.

<sup>5</sup> Dennett, 1991, p.430.

<sup>6</sup> Dennett refers to the self as a real abstraction in several places, but most notably in Chapter 13, 'The Reality of Selves' in *Consciousness Explained*, where he describes the self as a centre of narrative gravity and in 'Real Patterns,' *The Journal of Philosophy*, Vol. 87, 1991, pp.27-51.

<sup>7</sup> I have addressed his narrative self in Chapter 3. I engaged in a thorough critique of Dennett's account of the self in an earlier work. See *Dennett's conception of the self*: MA, Flinders University, 2002. Some of the following exposition and discussion will contain ideas drawn from that dissertation.

it is our capacity for *self*-conscious thinking that presupposes both self and consciousness. So Dennett has to give an account of self-consciousness. He does this first, by denying the sorts of things we think are true of the self and second, by explaining away the phenomenology which gives rise to the belief in the self. He then offers his narrative account of the self as an alternative to the Cartesian Self. Dennett's main claims against the existence of the self can be summarised as follows:

- (1) there is no inner thing in the brain, no central meaner or controller that constitutes a self;
- (2) the self is not permanent and unified but disparate and changeable over time;
- (3) we lack self-knowledge: most of what occurs in the brain is unknown to us and we don't know what we will do or say until we do or say it;.
- (4) there is no central locus or arena in the brain through which events must pass or be presented to qualify for the title of 'conscious.'

Dennett goes on to critique the idea of 'global workspaces' or central processing mechanisms as being covertly Cartesian.<sup>8</sup> As a consequence, he ends up positing a deflationary account of consciousness itself. If there is no central or single stream and there is no phenomenology, then consciousness becomes deplete of its defining properties. According to Dennett, 'consciousness is gappy and sparse and doesn't contain half of what people think is there.'<sup>9</sup> It is not a plenum. The 'richness' and continuity of our mental world is merely the richness and continuity of the physical world. This world is not represented in consciousness, even if that is the way it seems. As Dennett so wittily puts it, there is no MEdium, no special 'place' in the cognitive network which has to receive perceptual information in order for it to be conscious to 'me'.<sup>10</sup> There is no mind-pearl or soul-pearl that knows and understands. There is no real 'intrinsically-responsible' self. What we need to do, according to Dennett, is understand, '...naturalistically, the ways in which brains grow self-representations, thereby equipping the bodies they control with responsible selves when all goes well.'<sup>11</sup>

---

<sup>8</sup> See ie Baars, *Cognitive Theories of Consciousness*, Cambridge University Press, London, 1988 and Metzinger, 2003, in this chapter for examples of a central or global space.

<sup>9</sup> Dennett, 1991, p. 366.

<sup>10</sup> Daniel Dennett, 'The Message is: There is no *Medium*' in *Philosophy and Phenomenological Research*, Vol. 53, no. 4, Dec. 1993 pp. 919-931.

<sup>11</sup> Dennett, 1991, p.430.



### 5.3 Dennett's model of consciousness

Dennett begins his account of consciousness, mind and the self by putting forward a complex model of how intelligent-seeming behaviour might have arisen prior to the advent of an intelligent agent. He postulates that so-called 'intelligent behaviour' could have evolved over time by the joint actions of semi-autonomous, non-intelligent mechanisms. First, he sets out a very complex evolutionary model to show ways in which the random, non-discriminatory forces of natural selection could act on a simple organism to produce a more complex structure. He then illustrates how the same selection mechanism could select individual, relatively autonomous, relatively simple mechanisms that were adapted to taking care of specific functions or which could respond to very specific environmentally salient features in a non-conscious, non-intelligent way. These mechanisms, in concert, would produce successful coherent action.<sup>12</sup> He supports this hypothesis by using analogies drawn from computer technology. Advances in Artificial Intelligence and computer science show how purely physical systems can manifest supposedly human properties and powers without recourse to odd metaphysical entities like selves or agents – Shakey, SHRDLU and CADBLIND Mark II are some examples.<sup>13</sup> In reality, such systems are comprised of complexes of very simple subsystems which are specifically designed to 'take care of' localised pockets of specific types of information in a series of subroutines. The end result looks 'intelligent' but each mechanism is 'dumb' in itself.

According to Dennett, neither evolutionary theory nor computer technology needs to postulate the presence of an inner self to explain seemingly coherent, intelligent responses. In this way, Dennett puts forward a coherent argument to show that some forms of intelligent action can not only be explained by, but are also achievable from, non-intelligent (and non-organic) mechanisms.<sup>14</sup>

---

<sup>12</sup> This outline is necessarily brief and does not do justice to the immense detail in Dennett's model, prior to memetic evolution. However, the complex evolutionary path he posits adds weight to my later claim that the human biological self is far from minimal.

<sup>13</sup> These examples are illustrated in detail in Chapter 10, 'Show and Tell' pp. 285-297. He also discusses John Anderson's ACT\* and Rosenbloom, Laird and Newell's Soar programs in 'The Architecture of the Human Mind', *Consciousness Explained*, 1991, pp. 265-268.

<sup>14</sup> In this regard Dennett reflects standard contemporary thinking about the nature of the mind and its relationship to the brain. Similar accounts can be found in Owen Flanagan, *The Science of the Mind*

However, Dennett's task is to explain the *human* psyche in all its wonders. This will involve explaining how consciousness could have evolved from systems that were non-conscious. For this to work, consciousness can only be a late acquisition, superimposed onto non-conscious processing. In fact, Dennett argues that human consciousness is not a biological feature or a product of biological evolution. Rather, it is a by-product of the effects of social evolution, via meme-invasion. This makes the brain a necessary but not sufficient condition for (human) consciousness, mind and selfhood. For the latter, one needs information of the right kind. According to Dennett, the brain works as a highly complex, parallel and information-processing system in which localised, distributed specialists respond to and record content which then enters informational streams within the cognitive network. These streams of content are continually changing and evolving as new content pours in from various levels and places within the cognitive network, resulting in multiple drafts of related content. These streams of content can be tapped at any time by what Dennett refers to as 'probing' (either an external or internal stimulus that elicits a response from the system).<sup>15</sup> That response then becomes 'fixed' and is taken as evidence of the contents of consciousness at that particular time. This means consciousness is just the capacity to respond in some way when triggered or probed, by accessing whatever content-stream is available to report on or act on at that time. If probed later, it is likely the response would be different. Hence, one is conscious of whatever one says one is conscious of. There is no veridical account, no real fact of the matter about what one *really* was or wasn't conscious of. Neither is there *phenomenal* consciousness as such; all there is are perceptual judgements at various stages of development (drafts) and whatever enters memory as a result of probing at a particular time.

---

(2<sup>nd</sup> edition), MIT Press, Cambridge, 1991 and George Rey, *Contemporary Philosophy of Mind*, Blackwell Publishers, Mass., 1997. Even Searle assumes that A.I. is possible, although he claims the medium is relevant. John Searle, 'Minds, Brains and Programs,' *Behavioral and Brain Sciences*, Vol. 3, 1980, pp. 417-457. Debates focus on the various merits of either classical computational or connectionist models of cognition.

<sup>15</sup> Dennett defines consciousness as the capacity to respond to informational content in some way. The response is produced by probing, which is an environmental demand of some kind or an internal thought process. The response enters memory and becomes 'proof' of the contents of consciousness at that time. See *Consciousness Explained*, Chapter 5, 1991, pp. 134-138.

Dennett claims that a complex, non-centralised system of this type can explain human action without recourse to a Cartesian self or a Cartesian theatre. There doesn't need to be a particular place or a particular mechanism within the brain which does the work of the Cartesian homunculus. All one needs are various competing goals, localised mechanisms capable of responding to stimuli and streams of informational content, all which interact to produce seemingly unified actions. There are obvious advantages to this model of human action. There is no need to posit the existence of mysterious selves to explain action and it accords well with current scientific knowledge of how both the brain and artificial informational systems appear to work.<sup>16</sup>

However, as Dennett acknowledges, there is a huge gap between the actual mechanics of this type of system and the way we humans appear to operate, as well as what we take to be the subjective experience of being such a system. Mechanisms of the kind described above tend to produce rigid stimulus/response systems that demonstrate inflexible behaviour, which does not accord with the abundance of evidence of the richness and versatility of human behaviour. Humans often exhibit seemingly non-mechanistic, often innovative responses to novel situations. Likewise, multiple drafts of content fixation feeding into 'pandemonium-style contention scheduling' programs may produce seeming intelligent action but it doesn't explain the phenomenology. It doesn't capture the sense of being a self-conscious intentional agent capable of acting freely and thoughtfully who also appears to have rich, multi-dimensional phenomenal experiences of the world.

Dennett's solution is two-fold. First, he posits a further evolutionary phase, what he calls 'memetic evolution,' to explain our non-rigid and versatile behaviour. Memes or cultural units invade the brain via language or through other cultural symbols constructing a virtual, von-Neumannesque machine, which transforms the cognitive architecture such that it creates a mind with enhanced cognitive abilities. According to Dennett, it is by virtue of this machine that we are conscious and it is from this

---

<sup>16</sup> This refers to Ryle's 'ghost in the machine' or any similar assumptions in explanations of human actions.

consciousness that we posit a self<sup>17</sup>. Second, he claims that we are mistaken about our own experience of being a self-conscious agent.

In fact, our self-phenomenology is illusory. There is no ‘inner’ self and there is no phenomenal consciousness. What is going on in the brain and what we think is going on are two very different propositions. We have no incorrigible access to the contents of our minds, neither do we necessarily know what we will do or say until we do or say it. We learn about ourselves just as others do, by ‘seeing’ what we do or how we respond. In fact, the selves that we think we are, are a fictional creation, a narrative spun by various ‘word demons’ and mechanisms of content fixation around certain biographical details. As Dennett puts it,

Our human environment contains not just food and shelter, enemies to fight or flee, and conspecifics with whom to mate, but words, words, words. These words are potent elements of our environment that we readily incorporate, ingesting and extruding them, weaving them like spiderwebs into self-protective strings of narrative. Indeed...when we let in these words, these meme-vehicles, they tend to take over, creating us out of the raw materials they find in our brains.<sup>18</sup>

By introducing the idea of the memetically-constituted virtual machine programming the brain to produce a mind, consciousness and, hence, a self, Dennett attempts to accomplish several things. Firstly, he claims that the ‘machine’ overrides or adapts much of the basic hardware so that we no longer respond to stimuli by activating basic survival routines. The complexity, variety and sometimes, apparent pointlessness (i.e. no obvious survival function) of our actions can then be explained in terms of this transformation. Secondly, organisms whose brains are not susceptible to meme invasion cannot instantiate a meme machine. This means such creatures will not have minds, selves, consciousness and all the ensuing abilities because these are a direct result of 1) the changed architecture and 2) the informational content of the memes. Thirdly, it is intended to explain why we *claim* to experience the phenomenology in the way that we do and why it is all illusory.

What Dennett is actually saying is that, through the above process, human consciousness and the human self are socially constructed, both by the

---

<sup>17</sup> Dennett does appear to offer more than one explanation for consciousness, probing and meme-created software. It is the meme machine that constructs *a mind* with novel properties such as serial thinking and internal talking. This becomes illustrative of our conscious life when probed.

<sup>18</sup> Dennett, 1991, p. 417

transformation of the cognitive structure via the invasion of socio-cultural memes and by the socio-cultural content of the concepts themselves. This also means that access to a society or memes is essential for the development of a self-conscious mind, and that the type of memetic structure that gets established within the cognitive system will determine the abilities of that mind. As it is the contents of the mind which are representative of our consciousness, this in turn becomes representative of our self; what we are aware of about our self, whether real or fictional, is generated by the memes that fill our heads, creating ideas that we claim are our own. It is memes (not selves) that have agency, through an analogous process of natural selection. Memes survive because they are good at surviving, not because we think they are good memes.

The haven all memes depend on reaching is the human mind, but *a human mind is itself an artefact* created when memes restructure a human brain in order to make it a better habitat for memes.<sup>19</sup>

#### **5.4 Dennett's two selves**

Although Dennett claims that our self is a social construction, dependent on the instantiation of a meme-machine, he actually posits the existence of *two* distinct 'selves,' a minimal biological self and a personal 'selfy' self.<sup>20</sup> He claims that first we have a basic biological self composed of a variety of systems wired to differentiate itself from others, to recognise its own boundaries and to operate from its own self-interests. This biological self incorporates the biological mechanisms that maintain bodily stasis and the organism's integrity as an individual entity. Dennett uses the mechanism of natural selection to show how simple, blind, stupid processes, can respond (if somewhat crudely) to environmentally salient features in a productive but non-conscious way. They are not 'aware' of what they are doing; they act as they were selected to act. The end result, over time, is an increasingly complex system capable of complex responses to greater informational input. Although human behaviour may seem unique, it can be traced back to simpler mechanistic responses present in other life forms. Dennett draws analogies with various levels and varieties of organic life (such as spiders, dam-building beavers and termites) to

---

<sup>19</sup> Dennett, 1991, p.207 (my emphasis)

<sup>20</sup> Although Dennett talks about the 'minimal' self and 'narrative' self, his minimal self is not the same as Strawson's, a point I raise in relation to Gallagher in Chapter 6.

show that consciousness, and self-conscious autonomy, are not essential in producing complex responses or actions. The system is simply 'wired' to respond in certain pre-programmed ways by an increasingly complex set of interactions between a complex set of specialist sub-systems.

In the same pre-programmed way, the 'biological self' is wired to recognise the singularity of itself and what occurs to and within its borders. Dennett states that the basic minimal biological self begins with a recognition of boundaries. It has to differentiate itself from others, to know where it begins and finishes. Its organisation is thus 'self-interested'. It has a point of view in relation to itself and the world outside itself.<sup>21</sup> This 'birth of boundaries' creates a 'me' and 'the rest of the world'.<sup>22</sup> Thus we get a rudimentary self. Dennett then goes on to claim that this basic, minimal biological self, something that we share with the lowliest amoeba, is not a concrete thing, '...but just an abstraction, a principle of organisation,' although he does little to justify this assertion.

This minimal proclivity to distinguish self from other in order to protect oneself is the biological self, and even such a simple self is not a concrete thing but just an abstraction, a principle of organisation.<sup>23</sup>

This basic self is then transformed by complex socio-cultural units, mainly transmitted via language, to produce the human mind (and hence consciousness and the self). According to Dennett, we are also 'wired' to construct a personal self, a narrative fiction woven around various bits of biographical history and incorporating useful items from our social environment.<sup>24</sup> However, there is no single author and no definitive text; there is no 'I', no real nature. This self is spun by the various mechanisms or subpersonal agents in the cognitive network which interpret and process information about the self. This information is stored by a self-representation 'blip' (cognitive structure) designed to keep track of information about the self (but which, as Dennett is quick to point out, should not be mistaken for a self).

Neither the minimal biological self, nor the personal narrative self, is to be taken as a concrete entity; they do not exist independently of the effects of the mechanisms

---

<sup>21</sup> Dennett, 1991, pp.173-174.

<sup>22</sup> Dennett, 1991, p. 414.

<sup>23</sup> Dennett, 1991, p. 414,

<sup>24</sup> See Dennett's selfy-self, *Consciousness Explained*, 1991, pp.413-16.

whose concerted actions encourage us to posit their existence. I address this ontological position below. Despite this, Dennett is keen to draw a sharp distinction between the two, claiming that the biological self is *very minimal*, with hard-wired behavioural outputs that do little to explain the richness and diversity of human activities. It is the existence of the personal self that is of interest and which needs explaining. The personal self is the one we claim to be. Dennett posits that the individuality of each self, the richness of human endeavour, and the phenomenology that makes Cartesian selves seem so intuitively plausible, can all be explained by adverting to the changes wrought by memetic invasion. To this end, he proffers a non-Cartesian, non-reductive account of the self as a socially-constructed narrative. It is important to note that, according to Dennett, the existence of that narrative self cannot be explained by the actions of the ‘minimal’ biological self. It does not emerge from the actions of the biological self. It only emerges after the instantiation of the virtual meme machine, as does the accompanying phenomenology. This separates it conceptually from Gallagher’s and Bermudez’ core selves.<sup>25</sup>

Dennett argues that this account can explain why we have seemingly robust selves which not only appear continuous and permanent, but which also seem to be capable of meaningful decision-making. This, Dennett claims, is the result of the type of cognitive architecture which has been created by the meme machine. The resulting programme creates our phenomenology or the ‘user illusion’. For example, the ‘von Neumann’ machine processes information serially, creating the illusion of a single stream of consciousness; it installs the habit of self-stimulation by the advent of language acquisition and the ability to talk to oneself; it creates a seeming control centre, the place where all thought comes together. Not only that, but the very memes themselves determine how things will seem to us. So concepts like ‘self’ will affect the way we view our internal organisation because the word ‘self’ has certain defining properties in our language.

What humans also do when confronted by a talking, active system is posit a *centre* to the narrative, an author of the text. This doesn’t mean there is one, it just seems as

---

<sup>25</sup> See Shaun Gallagher, ‘Philosophical conceptions of the self: implications for cognitive science’ in *Trends in Cognitive Sciences*, Vol 4, no. 1, 2000, pp 14-21 and J. Bermudez, ‘Ecological perception and the notion of a non-conceptual point of view,’ in Bermudez, J.L, Marcel A., and Eilan N. (eds) *The body and the self*, MIT Press, Massachusetts, 1995.

though there is, just as there seems to be a ‘soul’ in the termite colony.<sup>26</sup> This third person positing of a self is what Dennett refers to as the ‘heterophenomenological self’, the centre of narrative gravity which is treated as though it were a self, the unifier of the body’s actions. So not only does it seem to the human subject that it has a self, it appears that way to everyone else as well. However, the ‘self’ as a real internal causal entity, with the capacity to take control and make decisions, the ‘self’ as a loveable character with a bad temper, this self does not exist; it is a fiction. And as Dennett claims, in line with Parfit;

[S]elves are not independently existing soul-pearls, but artefacts of the social processes that create us and, like other such artefacts, subject to sudden shifts of status. The only ‘momentum’ that accrues to the trajectory of a self, or a club, is the stability imparted to it by the web of beliefs that constitute it, and when those beliefs lapse, it lapses, either permanently or temporarily.<sup>27</sup>

Dennett wants to dismantle completely our phenomenologically-based belief in the existence of a self. If the phenomenology is shown to be illusory—if what it seems like to me is not, in fact, the case—then folk-theoretic claims about the self lose much of their credibility. Evidence from introspection becomes irrelevant in informing us about ourselves. For Dennett, the phenomenal properties we claim to experience as intrinsic to the system, like the sense of self, consciousness or a sense of autonomy, are actually illusory on several counts. First, there is no such thing as phenomenal consciousness per se; the system receives information from which it makes judgements about the world or oneself that are then proffered when required. We may claim that we feel pain or see red or that a curry tastes like this, but these are just examples of contentful states that have made it to memory or have influenced other states to precipitate action. Second, consciousness itself is a construct, emerging from the serial processing of the meme machine; it is not a state of being but merely the capacity to respond to stimuli in specific types of ways. How you respond reflects the ‘content’ of your consciousness but there is no ‘being conscious.’

Our experience of ‘being conscious’ is an illusion created by the concept of consciousness that exists in our historical period. Third, the changes wrought by the

---

<sup>26</sup> Dennett’s analogy, 1991, p. 416.

<sup>27</sup> Dennett, 1991, p. 423.



advent of memes create different abilities which did not exist in the system prior to that advent. So claims about one's self are merely claims about what you have become as a result of the memes which now reside in your brain. There is no deeper reality to the self. Dennett can now claim that selves don't exist in any tangible sense. All we have are posited centres of narrative gravity or heterophenomenological selves, and fictional, subjective accounts of the nature of those selves.

However, Dennett wants to claim that selves do exist in some way, even while denying them metaphysical reality. Selves, he claims, are 'real patterns of behaviour' and the narrative exists as a real, though fictional, construction in the same way Alice in Wonderland does.<sup>28</sup> Thus, it is not only useful to treat persons as though they are selves, as this helps explain and predict their behaviour, but to not treat persons as selves would be denying a *real* feature of the world of human interactions. This is one of the reasons why Dennett considers selves 'real abstractions' rather than just useful fictions.<sup>29</sup> As he says,

a self... is not any old mathematical point, but an abstraction defined by the myriad of attributions and interpretations (including self-attributions and self-interpretations) that have composed the biography of the living body whose Center of Narrative Gravity it is.<sup>30</sup>

## 5.5 Problems with Dennett's model of the self

Dennett's account is rather confusing. He makes seemingly contradictory ontological claims and puts forward several explanations for the same features (i.e. minds and consciousness). Dennett's case rests on the claim that minds, selves and consciousness are not inherent properties of human systems. In fact, if a human were outside a social structure and had no access to language or non-language-transmitted memes, then that human would fail to instantiate a mind, a self or consciousness.<sup>31</sup> It is by no means conclusive that this would be the case. The few instances of social or

---

<sup>28</sup> Dennett, end of Chapter 2, 1987.

<sup>29</sup> Dennett writes that he takes 'virtual machines and centres of gravity and, say, Australian democracy to be perfectly real, but also abstract.' Personal Correspondence, June, 1995, Appendix.

<sup>30</sup> Dennett, 1991, p.427, See Paul Ricoeur; *Time and Narrative*, University of Chicago Press, 1984 and *Oneself as Another*, University of Chicago Press, 1992 for earlier accounts of the narrative self.

<sup>31</sup> Dennett uses Sacks to substantiate these claims. See Chapter 2, *Seeing Voices*, Picador, London, 1991 where Sacks discusses the cognitive limitations of the deaf who do not acquire language (sign is a natural language). However, Sacks also claims that social isolation and deprivation are critical factors in inhibiting cognitive learning in deaf children. See extensive footnote p. 45.

linguistic deprivation we are aware of (such as the wolfboy of Aviron, Jeanie Doe or Sack's examples of deaf mutes), indicate no more than that extreme social and emotional deprivation is developmentally damaging and can lead to cognitive impairment.<sup>32</sup> They provide no evidence that the children involved had no minds or self-conscious awareness, nor that their impairments were the result of specific memetic deprivation.

The same argument is true of positing language as a meme vehicle. While one cannot deny that the lexicons of natural languages are learnt and that they are also culturally specific to some extent, it should be noted that language acquisition itself is a biological attribute, often following strict developmental pathways.<sup>33</sup> Its relationship to thought is part of an ongoing debate, as is the debate about the cultural specificity of linguistic concepts.<sup>34</sup> Both debates are outside the scope of this thesis. Nevertheless, I want to just mention that, even if language acquisition does influence our cognitive capacities, it may do so because of its formal, grammatical structure or syntax, rather than the memes it supposedly carries. According to Chomsky, although there are thousands of different languages, there are a limited number of variant grammatical forms which get repeated across cultures.<sup>35</sup> He also claims that there are distinct developmental patterns in the way children learn and use language, patterns which closely correspond to the age of the child rather than their socio-cultural circumstances. Chomsky uses this evidence to hypothesise that much of language competence is not socially acquired but innate. As he argues, time and available data are too limited to explain children's competency by positing an acquired system of rules.

There is no evidence to indicate that phenomenal consciousness is an illusion created by language or memetic acquisition, even if the *way* we perceive is influenced by

---

<sup>32</sup> See Steven Pinker, 'Baby Born Talking' in *The Language Instinct*, Penguin Books, London, 1994 and Oliver Sacks, 1991, p. 9.

<sup>33</sup> Just how much of language is innate and cross-cultural or culturally specific is still being debated. Chomsky argues that natural languages are heavily constrained by common basic grammars. However, his claim about the necessity of a generative grammar has recently been challenged by Daniel Everett's claims about the language of the Amazonian Piraha. See *Language*, Profile Books, 2012.

<sup>34</sup> See Dale Spender, 'Language and Reality; who made the world?' in *Man Made Language*, 1980, London, Routledge and Kegan; and J. Fodor, *Psychosemantics*; MIT Press, Mass., 1987 and Steven Pinker, Chapter 2 'Mentalese' in *The Language Instinct*, for an opposing view.

<sup>35</sup>Noam Chomsky, *Language and Mind*, Harcourt Brace Jovanovich, New York, 1972.

socially-transmitted beliefs. If Dennett's model is to work, if he is to show that the mind, the self and consciousness are meme-dependent, he has to show that human brains as they are biologically structured, *do not have the capacity for rationality, serial and future thinking or self-reflective thought* without meme invasion (traits that he identifies as representative of having a mind) and that these capacities, along with the contents of our thoughts, are purely *meme-determined*, hence culturally dependent. He doesn't do this successfully. There is no evidence to support the claim that humans who have limited access to language or access to only certain kinds or amounts of cultural memes lack consciousness, awareness of self or other mental properties reflective of having a mind. Even if such minds are different from what is considered the norm, they are still minds. As a counter argument, there is evidence of language profligacy in some cases of intellectual disability where there is no indication of any genuine understanding of the memes carried by the words or any concomitant increase in cognitive capacities as a result of the acquisition of the memes.<sup>36</sup> Minds or those capacities associated with minds seem to be the property of all functioning human brains, regardless of their exposure to memes. They are not just the prerogative of those who live in informationally-rich societies. Access to memes may explain some differences but will not explain the capacity to think serially or self-reflect.

Dennett's account of consciousness has been critiqued by many people over the years, mostly because of his use of meme invasion as an explanation for consciousness and the mind.<sup>37</sup> I have written a detailed critique of the meme machine in an earlier work, detailing why it can't account for our mind, self and consciousness and consequently, our phenomenology.<sup>38</sup> The gist of that argument was that the successful instantiation of a meme-machine seems to rely on a pre-existing conscious mind. As Block pointed out, and as mentioned in Chapter 3, it is highly unlikely that selves, minds and consciousness are good ideas first thought up by the ancient Greeks.<sup>39</sup> Also the very randomness of meme invasion, which Dennett

---

<sup>36</sup> Stephen Pinker argues that evidence of empty verbiage as in some cases of intellectual disability supports the independence of language and thought. See Chapter 1 in *The Language Instinct*.

<sup>37</sup> See Symposium: Dennett's *Consciousness Explained* in *Inquiry*, Vol. 36, 1992.

<sup>38</sup> See Y. S. Egege, *Dennett and the self*, MA Phil, 2002.

<sup>39</sup> Ned Block, 'Begging the question against phenomenal consciousness,' *Behavioral and Brain Sciences*, 1992, 15:2, pp. 205-206.

needs in order to explain non-survival and creative behaviour, makes getting the right mind-constructing memes into the brain a rather hit and miss affair. Yet most of us develop minds relatively easily, at quite young ages, and in all sorts of meme deplete and diverse situations. I also argued that Dennett under-rated the role of the biological self (which precedes the meme machine), in contributing to our sense of self. If consciousness is present in the minimal self, then he cannot claim that it is a product of the meme machine. Damasio and Block argue that consciousness may not be a higher order property but a fundamental aspect of phenomenal experience.<sup>40</sup> Dennett's 'probing' of multiple drafts could work equally well with the biological self and pre-memetic invasion.

Dennett's starting point is the implausibility of the Cartesian self, from which he argues that if there is no Cartesian 'soul-pearl' then there is no ontologically real self.<sup>41</sup> He goes on to argue that selves are constructions or useful posits like centres of gravity and electrons. Selves are not an intrinsic part of our biology or our cognitive architecture. They are neither singularities (non-reducible independently existing entities) nor are they reducible to more fundamental entities. A self is a linguistic term, a logical posit. It is what we say it is. This means that selves do not supervene on other states of the system. As Dennett says, selves are like any artefact; they have no ontological reality. They are an abstraction. He then insists they are a 'real' abstraction. As I indicated in Chapter 1, this kind of ontology is both confusing and restrictive. Dennett claims to use Reichenbach's distinction between abstracta, concreta and illata to show that selves are abstracta (real abstractions). In Reichenbach's framework, abstracta are not real, independently existing entities; they are constructions or 'reducible complexes.'<sup>42</sup>

Second, there are inferences to *abstracta*. These inferences are ... equivalences, not probability inferences. Consequently, the existence of abstracta is reducible to the existence of concreta. There is, therefore, no problem of their objective existence; their status depends on a convention... Third, there are inferences to other things which are not abstracta but which cannot become concreta either ... Our observations

---

<sup>40</sup> See Block, 'The higher order approach to consciousness is defunct,' *Analysis*, Vol 71, No 3, 2011

<sup>41</sup> Dennett, 1991, pp 28-9.

<sup>42</sup> Hans Reichenbach, 'The Projective Construction of the World' *Experience and Prediction*, Phoenix Edition, University of Chicago Press, 1961, p. 215.

of concrete things confer probability on the existence of *illata* [Inferred things] – nothing more...<sup>43</sup>

As can be seen from this quote, Reichenbach's abstracta are always *reducible to concreta*. They are equivalent to the concreta of which they are made. Reichenbach's example is a wall. This would mean that the selfy-self of the meme machine, the centre of narrative gravity, if it is to be considered a 'real' abstraction in this sense, should be reducible to concreta, those things that constitute it, in much the same way a wall is reducible to its bricks and mortar. However, according to Dennett, this is not the case with selves. They are not reducible to anything; they are an emerging 'pattern of behaviour' whose underlying causal structures bear little resemblance to its manifestation.<sup>44</sup> As Dennett is quick to claim, selves will not turn out to be accurate representations of internal cognitive states. No reduction of this kind is possible. We will not find the self molecules in the brain nor any other collection of cognitive mechanisms which will correspond to the self as we claim it to be. They are not straightforwardly reducible to specifiable concrete states even though our behaviour is explicable in terms of brain states of some kind and at some level.<sup>45</sup> Yet Reichenbach claims abstracta are *always* reducible to the concreta to which they are equivalent, unlike merely useful postulates.<sup>46</sup> If Dennett wants to claim that selves are 'real' abstractions, then he needs selves to be reducible to the concreta that generate them. Otherwise it is not clear why they are abstracta in Reichenbach's sense. However, Dennett insists that the self is a useful posit, nothing more. It is a story we tell about ourselves because we have a self-meme.

Dennett claims our selfy-self is not produced by underlying mechanisms but by the advent of the meme machine. Dennett further claims that the biological self itself 'is

---

<sup>43</sup> Hans Reichenbach, 1961, p. 215.

<sup>44</sup> Dennett's corollary is the game of artificial life, in which random on-off switches can appear to produce 'moving objects' when viewed from the right distance.

<sup>45</sup> It is important to note that there are lots of things which are not reducible in this way, ie to particle physics or to cellular biology. Properties like being a warm coat or a shady umbrella do not reduce to properties found in physics. (I owe this point to my supervisor, Dr Ian Ravenscroft) However, they still have causal powers and, despite the lack of reduction, are perfectly real. Dennett seems to acknowledge this but uses it to claim that these things are not *ontologically* real. This relates to Chalmers' point about objects in a commonsense ontology, Chapter 1.

<sup>46</sup> He says '... the nominalists were right in maintaining that the existence of abstracta is reducible to the existence of concreta.' Hans Reichenbach, 1961, p. 96

not a concrete thing but just an abstraction, a principle of organisation.<sup>47</sup> His reasons for claiming that the biological self is not a concrete entity are 1) it is composite; it is constructed from various 'dumb' mechanisms which take care of specific self-maintenance functions; 2) there is no single inner structure in the brain which could be considered a self (biological or otherwise); 3) the boundaries (of this self) are themselves porous and indefinite. He claims that,

the 'biological self is not like...the nucleus of a cell, or the gall bladder, or even the (distributed, poorly bounded) immune system. It is more like health or homeostasis...the biological self is not an organ, or even a single system, (like the immune system) though the proper functioning of such systems is what constitutes the biological self.'<sup>48</sup>

So the biological self does not qualify as a concrete thing even though the individual mechanisms of which it is composed can be said to exist as distinct physical entities. Dennett argues against its existence by claiming it is composite and lacks concise definition as a discrete individuated entity. One can talk about a (biological) self, based on the self-preserving properties of the organism and its apparent 'self-interested' actions. However, the term 'biological self' merely refers to the composite action of those individual, semi-autonomous systems responsible for preserving the integrity of the organism, hence its abstract status. This argument would seem to rule out many things that we would want to say existed – car engines, cities, universities, football teams, to name just a few.

In Reichenbach's ontology there are three categories - abstracta/concreta/illata; the abstracta are general terms for observable concrete particulars or posited entities which refer to complex interrelations between a complex of concreta. The illata are the posited unobservable causal entities, like electrons. According to Reichenbach, only the concreta and the illata have ontological status. Using this framework, Dennett claims that selves are neither concreta nor illata; they are abstracta. A reductive ontology like Reichenbach's, however, is problematic. A wall is abstract because it can be reduced to its concrete parts (bricks and mortar), but these can also be reduced to their concrete parts (sand, straw and concrete) making them abstracta. Ultimately, these can be reduced to smaller constituents which ultimately become illata, which are not concreta because unobservable. What counts as abstracta or

---

<sup>47</sup> Dennett, 1991, p. 414.

<sup>48</sup> Dennett, *Personal Correspondence*, Adelaide, June 1995, Appendix.

concreta will depend at which level of description one stops at. Reichenbach himself claimed this is a matter of convention or is arbitrary. Under this aegis, both Dennett's selves could qualify as either abstracta or concreta. However, Dennett would disagree. Dennett claims that the selfy-self is an abstraction because *it is not* reducible to concreta; he also claims the biological self is an abstraction because *it is* reducible to concreta. Neither position is viable. It not only appears that he is using different senses of the term 'abstraction' but he is applying different criteria to each 'self'.

Dennett's use of the term 'abstraction' for both the biological self and the selfy-self is quite deliberate, if theoretically flawed. As I outlined in Chapter 1, he starts from a sceptical position in relation to the self. Within Dennett's restrictive ontology, the Cartesian self can only be said to exist if it can be shown to be singular and this, for Dennett, implies a singularity, a self-neuron or soul-pearl. There is no such thing, ergo no self. The self that he puts forward is an epiphenomenal product of the changed cognitive architecture of the brain. No meme machine; no self. He can then argue that all the self phenomenology we claim to experience and which is at the heart of positing a self, is an illusion. If he were to then give concrete status to the biological self, one could argue that there is a *real self* at the heart of the system and there is a real basis to our phenomenology. Dennett wants to rule out that move. He does this by claiming that that self is also an abstraction. Abstractions are not fundamental causal parts of our world. In actual fact, he has no grounds to dismiss the biological self in this way.

## **5.6 Why Dennett can't argue against phenomenal consciousness**

A key reason why Dennett does not want to allow any kind of reality to accrue to the biological self is because he does not want this self to ground our phenomenology. If the biological self could be shown to create (some of) that phenomenology then 1) the meme machine would lose its strength as an explanatory tool; 2) the phenomenology would not be completely illusory; and 3) it would place phenomenal consciousness as an underlying property of the system rather than a late addition resulting from the construction of a meme machine. It would also lay the ground for a viable non-Cartesian self. Dennett claims that the self has a natural *disunity*,

reflective of a distributed or de-centralised system of localised, semi-autonomous information-processing mechanisms. The illusion of continuity, unity and even subjectivity are the unifying effects of the narrative. To justify these claims, Dennett spends a great deal of time dismantling qualia and our phenomenology. According to Dennett, there is no sense of oneself; there is just information about oneself as a physical system. It is non-phenomenal. There is nothing that it is like to eat cheese, just as there is no state of consciousness. The basic processing mechanisms of the brain are, he claims, non-conscious. According to Dennett, consciousness is a late developmental software addition, overlaid onto the non-conscious hardware. It is a capacity to access certain information and behave accordingly, rather than being a feeling or a state.

I argued against the narrative account of the self in Chapter 3, so I will not deal with this here. I also argued against the idea that we are not phenomenally conscious, as is presumed by most narrativists. Phenomenal consciousness, as a result of the narrative, can only be a late acquisition, not fundamental to the human (or other) system. I claimed this is a bizarre outcome of the narrative construction of the self. As mentioned in Chapter 3, some narrativists do suggest a core consciousness or self. If there is a basic or core self, there is likely to be a primitive subjective experience.

Because Dennett deflates consciousness to just ‘access,’ he denies that we can reliably know anything about our self from the ‘inside.’ Dennett denies the validity of introspection as a means of knowing or experiencing because he sees ‘mental states as constructs attributed to persons by outside interpreters.’<sup>49</sup> This hetero-phenomenological approach means any reliable knowledge of one’s mental states can only be gained from this objective perspective. Yet at the same time, Dennett claims that ‘...of all the things in the environment an active body must make mental models of, none is more crucial than the model the agent has of itself.’<sup>50</sup> He goes on to claim that the need for self-knowledge goes beyond the tracking of mere bodily movements. We need to know ‘our internal states, tendencies, decisions, strengths and weaknesses...’ It is difficult to know how we could have accurate knowledge of these things from the outside. Dennett’s own view on the indeterminacy of

---

<sup>49</sup> Hannan on Dennett, ‘Two Versions of Non-reductive Materialism’ in *Subjectivity and Reduction*, Westview Press, USA, 1994, p.79.

<sup>50</sup> Dennett, 1991, p.427.



interpretation means we could attribute several totally different sets of intentions to the same behaviour.<sup>51</sup> According to Dennett, there would be no fact of the matter about which set was right. This means that adopting the heterophenomenological stance is unlikely to produce an accurate portrayal of our mental life. Self-awareness or the capacity to ‘track’ from the inside would be a much more practical option.

There is growing evidence to support the idea that consciousness may be a basic property of the human cognitive system or cognitive systems in general. Bermudez argues for a form of non-conceptual self-consciousness that is rooted in our perceptual mechanisms at the bodily level of interaction with the environment.<sup>52</sup> He gives a very detailed account of a complex range of biological mechanisms which serve to individuate an active organism like a human being from its environment, as well as mechanisms which serve to identify it to itself.<sup>53</sup> He refers to the work of J.J. Gibson who provides a detailed account of the way visual information enters the visual system already delimited by self-structuring invariants.<sup>54</sup> Exterioception and interoception work equally to discriminate self from other and to provide that sense of unity and singularity that is self-defining. Primitive self-consciousness or awareness of oneself as an individual organism of a particular kind, could ground one’s self-concepts, as I have argued for in relation to Dennett’s minimal biological self. It could also provide an infallible basis for self-identification. In fact, if an account of this kind were right, one would only expect problems with the self-concept and the use of the indexical ‘I’ when there was a malfunction at the level of these biological mechanisms. As I suggested in Chapter 4, Cotard’s and other DRS seem to indicate just that. According to Damasio, conditions like Cotard’s represent a *loss* of phenomenal bodily consciousness indicating that the default position, when the appropriate mechanisms are working properly, is to experience ourselves as embodied. This would place consciousness at a primitive or foundational level and

---

<sup>51</sup> See Dennett’s comments pp. 46-49 in ‘Real Patterns’, *Journal of Philosophy*, vol. 87, pp.27-51, 1991.

<sup>52</sup> See Jose Luis Bermudez, Chapters 5 & 6. *The Paradox of Self-consciousness*, MIT Press, Bradford, 1998 and Bermudez et al (eds) *The body and the self*, MIT Press, Mass., 1998.

<sup>53</sup> Jose Luis Bermudez, ‘Ecological perception and the notion of a nonconceptual point of view’ in *The body and the self*, 1998.

<sup>54</sup> J.J. Gibson, ‘Foundations of Ecological Optics: Part I’ in *Reasons for Realism: Selected Essays of James J. Gibson*, (eds.) Edward Reed and Rebecca Jones, L. Erlbaum, Hillsdale, 1982.

would argue against Dennett's claim that self-awareness or our self-conception is a purely social phenomenon produced by memes.

Both of Dennett's two selves have grounds for being more than abstractions. However, this would undermine his thesis that selves are not real, no matter what they might look like. He also has limited grounds to claim we are not phenomenally conscious, a move that Metzinger also makes and which I address below. As a consequence, Dennett has to under-rate the role of the biological self and its impact on our self-phenomenology. However, the biological self is not as minimal as Dennett makes it seem. It has multiple individuating and self-identifying functions and could easily ground our sense of self. Bermudez and Damasio show that, at the least, there are grounds to think that the 'sense of self' is not an illusion or a product of social conditioning. As suggested in the preceding chapter, the phenomenology is strongly linked to the correct functioning of specific self mechanisms in the brain, that would be part of the biological self. This makes it less likely to be the result of a meme. There is also a case to be made for considering the minimal biological self a concrete entity, and a plausible candidate for selfhood.

## **5.7 Metzinger, phenomenal scepticism and positing the Phenomenal Self Model (PSM)**

Metzinger's model of the self has some interesting parallels with Dennett's version of the self as set out in *Consciousness Explained*. Like Dennett, it is a concerted attempt to explain our recalcitrant phenomenology within a physicalist framework. Given the difficulties of explaining self-consciousness and phenomenal experience using a (teleo) functionalist information-processing model, it is no surprise that his solution is also similar to Dennett's; he argues that there really is no such thing as the phenomenal subject. His dismissal of the self, like Dennett's, relies on dismantling our conception of phenomenal experience. It is critical for the feasibility of his model that he demonstrates the flaws in our every day 'folk' conceptions of perception and our belief in direct realism, which is at the heart of our belief in a self-conscious self. Dismantling our belief in direct realism will also undermine the reliability of our claims to genuine self-knowledge. Like Dennett, Metzinger believes that the Cartesian self is non-existent and that the self of experience is an illusion. To explain our sense of self he posits a representational model of the self, what he calls a

*phenomenal self model* (PSM). The PSM plays a similar role to Dennett's self-representation 'blip.' There are, however, some important differences between Dennett and Metzinger.

Metzinger's account is representative of a functionalist model of cognition that ends up negating the very phenomena it is trying to explain, such as qualia or self-consciousness.<sup>55</sup> The self and its phenomenology are reduced to the effects of a distributed data processing network which causes certain behaviours and creates certain illusions. In Metzinger's account, the phenomenal self is not an existing entity, neither is it a social or linguistic product. It is a simulation of a self that results from the brain synthesising phenomenal data as if from a singular perspective. This is then represented in the cognitive system which creates a PSM. Although it includes autobiographical details, it is not a purely narrative self like Dennett's. Metzinger's self-model is grounded in perceptual information and phenomenal experience. In fact, his PSM is, he claims, a distinct theoretical entity that will have a locatable neurological address; it can be 'found by suitable empirical research programs.'<sup>56</sup> However, just as Dennett warns against thinking his self-representation 'blip' is a self, Metzinger is equally emphatic about the status of the PSM. Although the *model* can be said to exist, what it models (the phenomenal self) does not.

Metzinger's explication of his PSM is highly complex and complicated. Each chapter is divided into a detailed exposition followed by the application of 12 multi-level constraints. As with Dennett, I cannot do justice to the wealth of detailed exposition and must leave out much of the neurological/technical detail. However, I will attempt to synthesise those aspects of his model that convey his main claims about the self, how it is constituted and how this is justified.<sup>57</sup> While I attempt to be as fair as possible and to do justice to what is a dense text, it should be noted that the text suffers from a lack of clarity.

---

<sup>55</sup> See Nicholas Maxwell, 'Three philosophical problems about consciousness and their possible resolution,' in *Open Journal of Philosophy*, Vol. 1, no. 1, 2011, pp. 1-10. DOI:10.4236/ojpp.2011.11001 for a critique of functionalism and the explanatory gap.

<sup>56</sup> Metzinger, 2003, Chapter 6, p.303 and p.411 '...something that will be *empirically* discovered...as a specific stage of the global neural dynamics in the human brain, characterized by a discrete and unitary functional role.'

<sup>57</sup> In his review, Josh Weisberg claims that Metzinger offers a 'sweeping and comprehensive tour through the entire landscape of consciousness studies' that lays out a 'rich and stimulating theory of the subjective mind.' *Journal of Consciousness Studies*, vol. 10, no.11, 2003, p 89.

Metzinger claims that his book is about ‘consciousness, the phenomenal self, and the first-person perspective.’<sup>58</sup> His overarching thesis is that selves do not exist, hence the title. He claims that what we have instead of a self is a *phenomenal self model* (PSM), which is what he claims is the ‘folk’ concept of self. In addition, he puts forward the idea of a *phenomenal model of the intentionality relation* (PMIR), which is the equivalent of our self-perspective.<sup>59</sup> He is very careful to clarify the point that the PSM is a *representation* of a self, not a self, and even the *phenomenal self* of the representation is not an entity but a process. While the PSM represents the self, it is the PMIR that does all the ‘self’ work. It is the active part of the model that takes the role of the agent. The PMIR produces action through ‘...an ongoing representational dynamics collapsing a phenomenal model of the practical intentionality relationship into a new transparent self-model.’<sup>60</sup> It is unclear exactly what this means or how it works, but he seems to be claiming that the PMIR is another kind of model that represents the subject-object relation. As I understand it, the PSM gives us a model of the self while the PMIR gives us a model of the self’s relation to the world (of objects) which creates its perspective. The PMIR allows the self (PSM) to be aware that its attention is directed towards an object, both externally and internally.

For Metzinger’s model to work, he has to first create an epistemologically sceptical framework for phenomenal experience. As with Dennett, this leaves him free to posit what seems intuitively implausible, i.e. that we are not really *self*-conscious, there is no subject and phenomenal experience is an illusion. As a consequence, there is no self to accommodate. He claims that phenomenal experience is ‘an on-line simulation or hallucination’ and that the phenomenal subject is ‘a functionally adequate but epistemologically unjustified representational fiction.’<sup>61</sup> Metzinger’s argument is as follows: if the phenomenal subject is the essential ground for believing in a self and it can be shown that there is no phenomenal subject, then there is no longer any ground for believing in such a self. Metzinger then posits the PSM as the only plausible ‘self’ that can be said to exist.

---

<sup>58</sup> Metzinger, 2003, p.1.

<sup>59</sup> See also T. Metzinger, and V. Gallese, ‘The emergence of a shared action ontology: Building blocks for a theory’, *Consciousness and Cognition* 12 (2003) 549–571.

<sup>60</sup> Metzinger and Gallese, 2003, pp. 563-4.

<sup>61</sup> Metzinger, 2003, p. 51 and p.58.

Metzinger justifies his scepticism by arguing that our subjective conscious experience rests on a temporal fiction. He claims that what we take as a direct, immediate experience of x is actually a representation of x. According to Metzinger, object identification is an ongoing process that creates the ‘illusion’ of a solid object. The phenomenal experience of an object is the end product of the activities of an *object emulator*.<sup>62</sup> This means that my supposedly direct experience of a blue-flower-at-this-moment is an illusion because it takes time for the cognitive system to record all the diverse and distributed sensory information and then to present it as a single direct perception of a ‘blue flower’ that is happening right now. Although it is represented as occurring ‘NOW’, this is false. Additionally, this complex percept is represented to the system as a conscious phenomenal experience happening to ‘me,’ another representation that is also temporally spread. Even though the experiencing subject is considered real, just as the experience of now is considered real, both are an illusion. There can be no direct perception of anything because it takes time to form a percept. He claims, ‘no such thing as absolute actuality exists on the level of real-world information flow in the brain.’<sup>63</sup> This means we can never be in touch with actual reality as it happens, the actual present. Thus, the phenomenal experience of ‘now’ must be a construct.

According to Metzinger, this constructed *hypothesis* results in ‘a simulated NOW.’<sup>64</sup> This has serious epistemological implications because any (or all) of our conscious representational content could always be false. They are not epistemically justified. A major implication of this position is that the claims that we make of ‘knowing’ x happened because we experienced x are no longer reliable because we are *never* phenomenologically present. It is a physical impossibility because of the way our sensory apparatus is wired to our brain. There will always be a time delay and we only have access to the end product or representation. Metzinger goes on to define mental representation as,

---

<sup>62</sup> Metzinger, 2003, p.23.

<sup>63</sup> Metzinger, 2003, p.26.

<sup>64</sup> Metzinger, 2003 p.57.

...a process, whose function *for* the system consists in representing actual physical reality within a certain, narrowly defined, temporal framework and with a sufficient degree of functionally adequate precision.<sup>65</sup>

According to Metzinger, all phenomenal experience is an indirect *representation* of reality that is always past. There is no possibility of direct experience of reality because there is always a time delay between the event to be cognised and the cognition. Any experience of direct realism is an illusion or an hallucination. In fact, Metzinger goes one step further and claims that phenomenal experience is really ‘a simulation of *virtual* reality, not reality itself (my italics).’<sup>66</sup> This makes it two steps removed from direct contact with reality.

Metzinger then introduces the concept of ‘autoepistemic closure’ to explain why it doesn’t seem this way to us cognising humans. Autoepistemic closure is the system’s blindspot. It prevents the illusion of real world access from ever becoming apparent.<sup>67</sup> From a first-person perspective, it always seems as if our experiences are both *of* the world and *in* the world, both of which are false. Metzinger terms this ‘the transparency of phenomenal representata’ which determines the structure of experience. As a consequence, the claims we make about our sense of self or our awareness/experience of things only tell us about the representata not reality. This means that Metzinger can claim that our phenomenology is false and can not be used as the rubric for defining the self. The closest thing we have to a self, he claims, is the PSM.

The content of the PSM is the content of the conscious self: your current bodily sensations, your present emotional situation, plus all the contents of your phenomenally experienced cognitive processing...one could even say you are the content of your PSM.<sup>68</sup>

## 5.8 How the PSM becomes a ‘self’

Metzinger is not the first to posit mental models in the brain that represent what is currently occurring in and to the cognitive system. Ramachandran posits the

---

<sup>65</sup> Metzinger, 2003, p.26.

<sup>66</sup> Metzinger, 2003, p.51.

<sup>67</sup> Metzinger, Chapter 6, 2003, p.131.

<sup>68</sup> Metzinger, 2003, p.298.

existence of potentially hundreds of different body maps, including of the sense of self that exist to help us survive.<sup>69</sup> Damasio also stresses the importance of body maps or models for keeping track of the state of one's body.<sup>70</sup> Metzinger draws from these ideas to construct his Self Model. He wants to use the idea of a model to illustrate the distinction between an actual self and a model of a self. Models are representations or simulations; they are not the thing itself. Metzinger then claims there is no real thing behind the model itself. The phenomenal self of the PSM is itself a construct that creates the illusion of selfness.

For all its wealth of detail, the mechanism/s by which the self or the PSM comes into existence are a little unclear and confusing. What follows is my attempt to extract the process of creating a working PSM. It seems to start with the phenomenal self of which the PSM is a model. As Metzinger is quick to point out, the phenomenal self is an illusion that begins with perception and the presentation of content. This content has no temporality; it is not 'represented' in the system because the human system lacks the internal resources for it to enter perceptual memory. These presentata are 'stimulus correlated states' that are objects for our attention, like colours, atmosphere, the state of our own body.<sup>71</sup> They are transient and just provide information 'that we are seeing' not what we are seeing. He uses blindsight patients to support the claim that one can have visual content without consciousness. At this point, there is no subjective conscious phenomenal experience. Phenomenal presentational content comes into play when the sensory data is available for cognition, attention and action control. Simple consciously experienced presentational content is what we would call sensory consciousness. To clarify Metzinger's point, Weisberg claims that while these mental states are conscious, they 'are not *for* anyone.'<sup>72</sup> They are not subjective because there is no subject. This seems to imply they are not from any point of view. Metzinger claims that this sensory information is ineffable, highly specific, contextual and transitory. It represents the NOW and is not recalled in memory. He claims that presentational

---

<sup>69</sup> V.S. Ramachandran and Sandra Blakeslee, *Phantoms in the Brain: Probing the Mysteries of the Human Mind*, Quill William Morrow, 1998. The Wilder Penfield body map depicts where specific body parts are represented on the brain's surface, illustrated as an homunculus with grotesquely distorted lips and hands to show the disproportionate sensors to these areas.

<sup>70</sup> Damasio, *Descartes' Error: Emotion, Reason and the Human Brain*, London, Papermac, 1996.

<sup>71</sup> Metzinger, 2003, p.88.

<sup>72</sup> Weisberg, 2003, p. 93.

content serves an important function because it discriminates reality (external world) from the internal world. However, even though it represents primitive, non-conceptual mental content it does contain elements of self-identification. I return to this point later

Metzinger claims that presentational content cannot have properties because properties are cognitive constructs. At the same time, his ‘simple’ presentation or conscious phenomenal presence is not so simple as it first appears. As he has said earlier, perception unfolds over time, which is why he warns against undifferentiated models of perception.<sup>73</sup> Conscious experience is a ‘remembered present.’<sup>74</sup> He also acknowledges Dennett’s cautionary note about presentational/ representational content being covertly Cartesian and needing an internal witness.<sup>75</sup> This is why he claims it is property-less in its primitive form. Yet Metzinger puts forward a highly representational account. He has presentational content that gets re-presented as phenomenal experience or a phenomenal self and this representational content then gets re-presented into a simulation of a self-model. It is not clear why it has to be constantly re-presented if there is no ‘place’ in the system where it needs to be ‘presented’ to be conscious. He also claims a *phenomenal* experience is just one which contains phenomenal information. It counts as a *conscious* experience if that information becomes globally available. By the same token, a representation becomes a *conscious* representation if it becomes globally available to the system. For Metzinger, a self-conscious experience means that the information relates to oneself, by contributing to one’s self-representational model.<sup>76</sup> However, adding in a globality constraint makes it seem even more like Dennett’s ‘closet witness,’ the Cartesian self inside the system that needs to view everything first.

While Dennett also posits a form of access consciousness, he resists positing a point that must be reached for consciousness to occur; he is also cautious of re-presenting as a matter of course. Metzinger never quite explains what it means to be globally

---

<sup>73</sup> Metzinger, 2003, p. 93.

<sup>74</sup> Metzinger, 2003, p. 98.

<sup>75</sup> Dennett, ‘There is no single, definitive “stream of consciousness,” because there is no central Headquarters, no Cartesian Theater (sic) where “it all comes together” for the perusal of the Central Meander.’, 1991, p.254.

<sup>76</sup> Metzinger, 2003, pp. 36-38.



available or how this is achieved. Instead, he works in reverse. If something is deemed conscious then it must be globally available.

According to Metzinger, there are many self-models within the human cognitive system that represent different facets of the human being. However, they are not conscious and, as a consequence, are not automatically part of the PSM. Again, he claims the models become conscious and part of the PSM by becoming globally available. Global availability allows the system to think or act on the information contained in the self-model. Metzinger claims that goal-directed actions need a ‘conscious self-model to *deliberately* initiate those actions.’<sup>77</sup> For the PSM to be a coherent self model, the disparate collection of phenomenal representations need to be united in some way. Metzinger bundles them together by a ‘higher order property’ of ‘mineness.’<sup>78</sup> This is like the property of ownership or belonging which links the self-states/self models and separates them from other non-self representations, similar to James’ mineness property that tied Hume’s bundle together.<sup>79</sup> Together, says Metzinger, they create the phenomenal self which is ‘the content of self-consciousness, given in phenomenal experience.’ Metzinger says this creates the property of phenomenal selfhood, normally termed ‘the sense of self.’ This phenomenal self is then represented by the PSM, creating the first-person perspective by representing itself to itself as a self. Metzinger states that,

The existence of a coherent PSM generates a pre-attentive self-world border and thereby realises the central condition for the development of genuine *introspection* (his italics), on the representational side as well as the phenomenological level of description.<sup>80</sup>

He claims that the presence of a (coherent) self-representational model introduces a self/world boundary *for the first time*. This is the beginning of subjectivity and objectivity. The PMIR then comes into play to provide the direction of attention, the self-perspective and the capacity for action. He stresses that the PSM is a representation of ‘virtual reality,’ not of reality. The content of a ‘conscious self-model of the body’ is constituted by a vast array of proprioceptive, tactile, and

---

<sup>77</sup> Metzinger, p. 299, 2003.

<sup>78</sup> Metzinger, p. 302, 2003.

<sup>79</sup> W. James, *The Principles of Psychology*, Chapter X, ‘The Conscious Self’, p. 361, 1890. Accessed July, 2010, <http://psychclassics.yorku.ca/James/Principles/prin10.htm>.

<sup>80</sup> Metzinger, p. 307, 2003.

vestibular bodily sensations. However, and contrary to what it seems like, he claims one is never in contact with one's body. This is because 'you' are the contents of an image of a body, which includes self-simulated contents, biographical memories and self-representational information. In fact, a self-model is 'a model of the very representational system that is currently activated.'<sup>81</sup> Mineness is a higher order phenomenal property of phenomenal content that is introspectively available.<sup>82</sup> By this he means that the property of it being 'about you' is globally available such that it can be part of the PSM.

## **5.9 Problems within Metzinger's model : Phenomenology and embodiment**

Metzinger's dismissal of direct realism and phenomenal experience is similar to Dennett's as set out in the preceding section. Perception is the end product of the actions of a temporally and physically distributed information-processing system, such as the visual system. The object of perception is a construct not directly perceived or experienced. Dennett calls this a 'judgement.' Metzinger calls it a 'simulation.' The reason they are 'simulations' rather than representations is because they simulate actuality, the belief that the experience is happening NOW. This belief is false because humans can never experience '...the simple fact that the content of their subjective experiences always is counterfactual content, because it rests on a temporal fiction.'<sup>83</sup> He goes on to make a further distinction by claiming that '[w]hen the intentional content of a simulation is represented as "temporally external" it will be experienced as a simulation, when not it will be experienced as a representation.'<sup>84</sup> This claim is a little confusing because, even though we can cognitively acknowledge the existence of simulations and representations, we don't experience them as such unless consciously engaging in a simulation. Nevertheless, there does seem to be scientific evidence to the effect that perception itself is a temporally spread activity. We also seem to be phenomenally blind to this process

---

<sup>81</sup> Metzinger, 2003, p. 302.

<sup>82</sup> Metzinger has 4 kinds of introspection that relate to phenomenal content and qualia. Each has a different level of accessibility leading to conscious experience of that content.

<sup>83</sup> Metzinger, 2003, p. 57.

<sup>84</sup> Metzinger, 2003, p. 59.

most of the time.<sup>85</sup> Despite this, I am not sure it has the sceptical implications that Metzinger draws.

Descartes also put forward a highly representational model of perception. Despite this, he argued that we could not doubt our conscious experience, regardless of how it was produced, even if we could doubt the content of the experience. This is because there has to be a subject who has the experience, no matter how false the experience. 'I' still experience a blue flower or a pain, even if both could be called either a simulation or a representation. Like Metzinger's use of the brain in the vat, Descartes' evil demon could not create a false sense of subjectivity, the experience of being someone having an experience of something. One could argue that, rather than being illusory, this is what it is to be conscious or to experience the world. Consciousness is experienced as a subjective state, which is why the zombie problem arises. Even if what I am experiencing is false, it does not negate the experience, or that 'I' am experiencing. This was Descartes' point in postulating the evil demon. Hence, one's phenomenal experiences are what they are. One doesn't need those experiences to conform 100% to 'objective' reality. As Metzinger himself claims, reality gets represented 'with a sufficient degree of functionally adequate precision.' Even if I am wrong about something, that doesn't negate my phenomenal experience nor my existence as a conscious being.

Metzinger pushes his point further by referring to the temporal delay. It is not just that the experience is a construction but that it is based on a temporal fallacy. It is not happening now; it is a remembered present. This makes it both a representation of an event and indirect. You are *never* in direct contact with reality. This means your sense of direct realism and of being in touch, 'being in the world' is an illusion. Any claims you make are not epistemically reliable because they could always be false. Consequently, your claims about yourself and your own subjectivity are equally fallible. In fact, they are false. There is no self, only a self model. Again, I am not convinced that Metzinger can draw such strong sceptical conclusions using

---

<sup>85</sup> Exceptions seem to be under some neurological conditions or when perceptual apparatus is damaged. The underlying process in these instances seem to be potentially opaque.

supposedly objective data or Dennett's third person, heterophenomenological approach.<sup>86</sup>

It seems to me that one's 'now' is when one experiences it. It is not when an objective outsider claims one should be experiencing it. Whether or not it is delayed makes little difference to me in that my experience of it doesn't change. Metzinger claims this is because the process is transparent due to auto-epistemic closure and that this is enough to discount the veridicality of perception and all self-knowledge. But it seems to me that this is what perception is; this is what it is to perceive something. My experience of things is at the time of my experience of them and this constitutes my 'now.' It doesn't seem to matter that my pain is 0.2msec after the stimulus reaches a point in my brain. The delay does not mean I did not really experience pain but only a simulation of pain. I am not convinced that the delay in registering something as a conscious experience leads to the conclusion that there is no phenomenal consciousness or that there is only a simulated object and a simulated subject.

There also seem to be grounds to query the extent of his phenomenal scepticism. We know from laboratory experiments that it takes time for stimuli to reach the threshold of consciousness. If the stimuli are too quick, they do not register in consciousness. This means that events have to persist in time for them to register as conscious phenomena. Objects and actions must, therefore, have some duration if we are to experience them, interact with them or enact them, especially if this is over long periods of time. This also means that much of what we experience can be said to be based on what is continually present right now during this temporal period because the stimulus has to be present long enough for us to register it as an experience. Objects do not stop interacting with the visual system once a perception has been recorded. That visual information is continually available to the system as long as the object is visually present. So one could equally argue that we are aware of what is happening as it happens. We are aware of the NOW.

---

<sup>86</sup> Ramachandran also claims that pains, experiences and the self are illusions, based on similar neurological evidence... He says 'your own body is a phantom, one that your brain has temporarily constructed purely for convenience.' V.S. Ramachandran & Sandra Blakeslee, *Phantoms in the brain*, New York: W. Morrow, 1998, p. 62. This seems an extreme claim. Despite this scepticism, Ramachandran doesn't deny consciousness. And the phenomenology still needs to be addressed. The 'objective' data does not always explain why x feels like x or why it feels like anything. Claiming the feel of x is just an illusion does not answer the question. It doesn't make torture less painful.

While it may be that some of our cognitive processes are simulations or representations, as Metzinger suggests, it is by no means obvious that *all* of our experiences and sensations can be described this way, or even need to be described this way. Metzinger claims that even pains and other bodily sensations are not directly experienced. Yet pain and other physical bodily sensations appear more direct. They seem to become conscious once the stimuli is presented, with no need of re-presentation. Representation may come later with the cognitive awareness of what it is a pain of, or where the pain is. This may require an adjustment of the bodily map. However, it is not obvious that there needs the level of representation that Metzinger indicates, particularly in the initial stages of pain or other sensory experience. Take the eating of an apple, for example. This is an action that takes time to complete and has multiple sensory effects, not least on the digestive system. According to Metzinger, it is representation all the way down. Not only is the sensation of eating the apple an hallucination but so is the sensation of the object in the hand and gut, the apple itself. The body eating the apple is as much a simulation as the self that is experiencing the whole process.

Yet the interaction of the object with the body need not be that indirect; the information can be just *presented* to the body. There is a real object, not just a simulated object. For Metzinger presentational content has to be pre-conscious because conscious content is representational. Many phenomenologists would argue that phenomenal experience is also conscious, even self-conscious because it is what constitutes the self by being self-defining.<sup>87</sup> Metzinger himself intimates as much when he claims that presentational content has some self-defining features. However, for his PSM to work, consciousness cannot be a basic property of the system or tied to simple perception. In his model, phenomenal consciousness is not subjective because it precedes the PSM. His model also seems to rely on there being *some thing* or *some threshold* within that requires information to be gathered together in a particular form for it to be recognised as an experience of something. If there is no self anywhere, it is unclear why this needs to be so. Again, this looks like Dennett's inner witness.

---

<sup>87</sup> Merleau-Ponty is the best-known proponent of this view but similar accounts can be found in Sartre and Nietzsche. See *Phenomenology of Perception*, London: Routledge & Kegan Paul, 1962. Zahavi raises a similar criticism in 'Being someone,' *PSYCHE*, vol 11, no 5, 2005. <http://psyche.cs.monash.edu.au/>

There are other inconsistencies in Metzinger. The validity of Metzinger's model rests on his dismantling of phenomenal consciousness and, thereby, dismantling the reliability of subsequent self-phenomenological reports. Metzinger spends the early part of his book dismantling the reality of perception on one hand and the cognising subject on the other to prove his point. He then engages in some rather odd sleight of hand. He re-introduces the phenomenal subject *as if* there were such a reality. He not only commits a whole chapter to discussing this phenomenology but alludes to it throughout his account *as if it were so*. He consistently refers to the experience of embodiment and phenomenal states as intrinsic to the human condition. According to Metzinger, having a sense of 'now' or being present is the hallmark of consciousness. Without this sense of immediacy, there would be no phenomenal states, no person.<sup>88</sup> He goes on to claim that 'the phenomenal self is always embedded in a phenomenal world, seamlessly and pre-attentively. Being self-conscious is being-in-the-world.'<sup>89</sup> This sense of immediacy and embodiment is reinforced by the 'invariant background of self-presence' such as bodily experience and somatosensory perception.<sup>90</sup> He talks a lot about the 'presentational constraint,' indicating that these are the kinds of factors that need to be taken into account in order for any theory of the self to be viable.

He claims that presentational content is the 'Now' and is distinct from simulation, cognition and memory. This 'phenomenology of presence' is a 'dynamical continuum' that would lead to a sense of temporality and continuity.<sup>91</sup> He insists that 'conscious content is always experienced as 'NOW'' and is always 'being present *as a self*.'<sup>92</sup> This is because it 'owns' the process. It is not clear what this latter phrase means, given there is no self that can own the process. But, again, this is how Metzinger talks. He makes it seem as though he were talking about the self of folk psychology with all the favourite trappings, particularly when it comes to the experience of embodiment. Yet Metzinger claims that if we were to get outside the self-model, there would be no phenomenal self to experience what it was like. We would be a system without a conscious self.

---

<sup>88</sup> Metzinger, 2003, p. 126.

<sup>89</sup> Metzinger, 2003, p. 307.

<sup>90</sup> Metzinger, 2003, p. 312.

<sup>91</sup> Metzinger, 2003, p. 126.

<sup>92</sup> Metzinger, 2003, p. 310.

Metzinger's emphasis on the importance of embodiment, phenomenal experience and a conscious self makes his account seem closer to our lived reality than it really is. He intimates that the 'experience' is both real *and* illusory; real because we cannot get outside the model to see the strings, illusory because it is created by the self-model. He acknowledges the importance of being a conscious self, while claiming that it is a fake. In a similar vein, Metzinger claims that '[a]ctions are first-person phenomena, and they are carried out by a conscious self.'<sup>93</sup> This would imply that the phenomenal self has to be aware of and direct the action it performs. It, thus, would be conscious of that action as it performed that action. It also means there has to be something like a conscious self that has the capacity to act. But this again seems to contradict Metzinger's claims about the lack of reality of such a self. In Metzinger's model, the system could operate just as well without the phenomenology. It doesn't seem to serve a purpose. The same goes for the self. If the self is pure illusion, then how does it enact anything and why is it so important to have a conscious agent with all its full-blown phenomenology?

Another example of Metzinger's inconsistency is when he claims that 'possession of a phenomenal self (not self model) is a necessary precondition for cognition and not its product.'<sup>94</sup> He refers to Damasio and his idea of a core self as being a necessary primitive form of self and self-consciousness that is 'given' in experience. This would seem to undermine the grounds for Metzinger's own PSM. He claims that phenomenal experience is both being present and being present as a self, reflecting its mutually constitutive nature.<sup>95</sup> According to Metzinger, a self-model serves an important function for the system, wherein '...a subjectively experienced, numerical identity of the self can emerge from a high degree of internal coherence.'<sup>96</sup> He then goes on to say that the experience of 'oneness' and dynamic fluidity may very well be features of our biology or reflect 'objective physical properties of our body' because the phenomenal content of our internal world supervenes entirely on 'synchronous and internal properties of our body.'<sup>97</sup> This seems to suggest that

---

<sup>93</sup> T. Metzinger and V. Gallese, 'The emergence of a shared action ontology: Building blocks for a theory', *Consciousness and Cognition* 12, p.561, 2003.

<sup>94</sup> Metzinger, 2003, p. 321.

<sup>95</sup> Metzinger, 2003, p. 311.

<sup>96</sup> Metzinger, 2003, p. 315.

<sup>97</sup> Metzinger, 2003, p. 322.

features of the internal subject mirror real features of the body. Thus, what I experience as a singular self is, in fact, true of my body (me). This could be interpreted as a form of privileged access to internal states, or at a minimum a mapping/matching of experience onto reality. This doesn't seem to match his strong claims of the illusory basis of the PSM and the phenomenal self on which it rests.

### **5.10 Problem with Metzinger: epistemic scepticism and ontology**

Metzinger's account of the self/ no self has broader sceptical implications. Cognising humans claim to have direct sensory contact with the world and privileged access to their own inner feelings. Our experience seems to be in the present, immediate and direct. Metzinger argues that all our phenomenal experience is an illusion based on a temporal fallacy. There is a time delay between the stimulus of the object and the response of the system to represent that object. Thus, our phenomenology is not epistemically reliable because it could always be wrong. This includes our belief in the phenomenal subject or a conscious self. Thus, we cannot have reliable self-knowledge because we cannot get beyond the blindspot of autoepistemic closure and the transparency of the PSM. This casts doubt on the reliability of any of our claims about the existence of a self and what we experience. Yet, our claims are mostly reliable, a point even Metzinger acknowledges. This means we could use a system of checks and balances to cross-reference our beliefs to make them more reliable. Likewise with self-knowledge. Metzinger goes further and even speculates about the possibility of us losing transparency and seeing that what we are calling our 'hands' is actually a misrepresentation, in line with Ramachandran's claim that our body is an illusion. So our PSM is 'not a good instrument for gaining self-knowledge.'<sup>98</sup> Bizarrely, what he seems to be saying is that we would see our 'hand' for what it really was. This could only be a representation of a hand, rather than an actual hand. It is hard to know what this would look like. Whatever it turned out to be, there would have to be some thing there; it wouldn't turn out to be a 'no thing.'

Dennett, of course, used a similar argument to dismantle the traditional claims about privileged access to internal states and immunity to error through misidentification,

---

<sup>98</sup> Metzinger, 2003, p. 335.



which I addressed in Chapter 4.<sup>99</sup> He also used representation and spatio-temporally distributed information-processing systems to throw doubt on direct realism. Metzinger's temporal delay argument plays a similar role. While one may appear to be phenomenally conscious, one is only conscious of oneself as a perceiving subject once the PSM is in place. That consciousness is not phenomenal consciousness or 'what it is like' consciousness, but global availability or access consciousness. The PSM, like Dennett's meme machine, creates the illusion of a cognising conscious subject or self. Consciousness is just the capacity to act on information or to have assertoric thoughts.<sup>100</sup> Metzinger and Dennett both hold that consciousness and selfhood are higher order properties that overlay a non-conscious process.<sup>101</sup> They both suggest that consciousness is just the capacity to say, think or do things when triggered by environmental features.<sup>102</sup> The end result is that phenomenal experience becomes an unreliable means of gaining knowledge of either oneself or other objects in the world. Thus, Metzinger's account leads to scepticism of phenomenal consciousness in the same way Dennett's does, despite his 'lip-service' support of the phenomenology and his claims that he is describing a conscious system.

Such epistemic scepticism will have an impact on all our knowledge claims because it is how we perceive the world and form judgements about the world. If we are unable to get beyond the transparency of autoepistemic closure, then it makes it difficult to see how we can rely on any of the claims we make, including those about the external world. If we should be sceptical of gaining self-knowledge, then we should be sceptical of external knowledge claims too. Both Dennett and Metzinger claim we use the same tools for gaining knowledge of ourself as we do for gaining knowledge of the world. There is no privileged access. If there is no privileged

---

<sup>99</sup> Dennett, 1991, p.428.

<sup>100</sup> This issue is too big to do justice to here but Block gives a clear expression to the problems with the representational, higher order thought (HOT) concept of consciousness. Assertoric thoughts seem similar to Dennett's 'probing' mechanism in that it is just the capacity to assert something. N Block, 'The higher order approach to consciousness is defunct,' *Analysis*, Vol 71, No 3, 2011.

<sup>101</sup> For similar views see J Weisberg, 'Misrepresenting consciousness,' *Philosophical Studies*, Springer, 2010; D Rosenthal, 'Higher order theories of consciousness,' *Oxford Handbook of the Philosophy of Mind*, eds B McLaughlin and A Beckerman, Oxford, Clarendon Press, 2009, pp. 239-252.

<sup>102</sup> Zahavi and Parnas are critical of highly representational models because they claim they lead to an infinite regress. They query how a thought becomes conscious by being the object of another thought and can't see how you get first-person givenness in this way. D Zahavi and J Parnas, 'Phenomenal consciousness and self-awareness: a phenomenological critique of representational theory,' in *Models of the Self*, editors S Gallagher and J Shear, Imprint Academic, U.K., 1999.

access, all claims are equally reliant on the reliability of our perceptual mechanisms and reasoning skills to come to useful conclusions. If one counts as genuine knowledge, so should the other, and vice versa.

If we accept the sceptical approach, knowledge of ourselves fares no better or worse than our knowledge of anything else. It is just as flawed or just as reliable. All this means is that the self stands in the same relation to veridicality as any other object in the world, no more but certainly no less. Despite this, we have managed to establish reasonably reliable bodies of knowledge in a vast range of areas. In fact, Metzinger applies a vast body of such knowledge to justify his account of the PSM. We should either be just as cynical of that body of knowledge as Metzinger says we should be of the claims we make about ourselves, or we should apply the same rigour to both. There is no reason why we couldn't develop a reasonable body of knowledge about our own phenomenal selves by applying similar standards of rigour. As we know from our pursuits of knowledge, it does not follow that x does not exist just in case some of the things we claim about x turn out not to be true. The self does not *not exist* because it does not have one or some of the properties of the Cartesian self.

### **5.11 Why can't the PSM be an alternative to the Cartesian self?**

Ultimately, Metzinger is putting forward a self model theory (SMT). He is quite adamant that his PSM is not a self and can't count as a self, even though we are the contents of the PSM. This is because it is a model of the phenomenal self that is also not a real self. It is the sum of the representations of self phenomena that is presented as a self by the PSM. Unlike the (folk) self, Metzinger claims that his PSM is a *distinct* (his italics) theoretical entity that can be empirically verified.<sup>103</sup> This makes it real in that it is (hypothetically) detectable. While the (phenomenal) self is essential in giving the system 'centredness and perspectivalness,' it is not a real self because it is based on an illusion.<sup>104</sup> The self that we experience is a presentation; it does not have the Cartesian properties of individuality, substantiality or

---

<sup>103</sup> Metzinger, 2003, p. 303.

<sup>104</sup> Metzinger, 2003, p. 303.

essentiality.<sup>105</sup> Metzinger concludes by saying we need to take a sceptical stance towards the self in order to move forward and get beyond the limitations of our phenomenological constraints.

I am not convinced that Metzinger presents strong enough grounds to dismiss the existence of the self. Any consciously experienced phenomenon is likely to be subserved by some kind of neurological or biological mechanism/s. The self would be no different, so on its own it is not a dismissal of the self, any more than it is a dismissal of vision or the mind or the immune system. I address this point in the final part of this chapter. One could even argue that Metzinger's PSM could count as a genuine self, despite its non-selflike description. It appears to be responsible for producing both our sense of self and the experience of the self as a distinct conscious subject. It is an entity in its own right and has, potentially, a temporal/spatial location. Metzinger dismisses this option because of its illusory core. There really is a self model, but the self it is a model of does not exist. As mentioned earlier, Strawson would claim that this move is illegitimate. His phenomenological constraints state that if there is something that subserves the self model that is an accurate representation of it, then that something counts as a self. I address this constraint in more detail in the following chapter.

Metzinger's dismissal of the folk self seems to rest on traditional anti-Cartesian grounds. As Zahavi points out, the self Metzinger is arguing against is a narrowly conceived Cartesian self, '...a mysteriously unchanging essence, a process-independent ontological substance that could exist all by itself, i.e., in isolation from the rest of the world.'<sup>106</sup> This narrow ontological position is reflective of Dennett's insistence on a singular 'soul-pearl' as the self of the folk. There are several other models to choose from. Metzinger's own PSM could be one such model. Zahavi refers to other existing self-models that recognise ipseity or a subjective 'givenness' of perception as a grounding for a pre-conceptual self.<sup>107</sup>

---

<sup>105</sup> Metzinger says that 'individuality (in terms of simplicity and indivisibility), substantiality (in terms of ontological autonomy) and essentiality (in terms of transtemporal sameness) are not properties of selves at all.' Metzinger and Gallese, 2003, p.626.

<sup>106</sup> Zahavi, 2005, p.8.

<sup>107</sup> Zahavi, 2005, p. 9.

Metzinger's model raises some interesting points about what would count as a self. I don't think the PSM does count as the (folk) self of experience. It negates the very phenomenology that grounds the idea of the self. There is no 'being self-conscious.' The phenomenal self could qualify but Metzinger would need to place consciousness as a more primitive self-defining property. His PSM is a late addition, higher order concept that does not adequately explain the robustness of early developmental experiences of selfness. It may rule out consciousness in young children and other animals. Like Dennett, Metzinger by-passes the need for explanation by claiming there is nothing to explain. We are in the Matrix or the equivalent of brains in a vat. This is claiming too much. Even if our knowledge is indirect and based on inference, one could argue that some of our claims about ourselves would be true. Even if the self were a process, it could still be considered real or existent. It doesn't need substantiality, but it would need unity and differentiation, properties that Metzinger claims are part of our biological structure.

## **5.12 Metzinger and phenomenology**

As Strawson states so clearly, it is because we have a sense of self that there is a problem about the existence of the self. The phenomenology is at the centre of issues about the self. This means it must be adequately addressed and accommodated. Metzinger deals with the phenomenology by attempting to show that it is an illusion created by the PSM. This means he is no longer constrained by that phenomenology in his explanation. Despite this, Metzinger does acknowledge the robustness of that phenomenology and its importance in making us feel embodied and a part of the world. He even claims the experience of a self, not just the idea of a self, is essential to our lives; it plays a functional role in decision-making and action; and it is impossible not to experience things from a position of selfness. Yet he dismisses it all as illusion and hallucination. There is no sense of being. This is an illusion created by the PSM and the transparency created by autoepistemic closure. Yet the poignancy of conscious life so aptly captured by Kim does not seem to be fully explained by this model:

We would regard a life as impoverished and not fully satisfying if it never included experiences of things like the smell of the sea in a cool morning

breeze, the lambent play of sunlight on brilliant autumn foliage, the fragrance of a field of lavender in bloom...'<sup>108</sup>

In sharp contrast to this dismissal of the reality of our phenomenology, Metzinger then uses the *phenomenological* reports of particular subjects as evidence to support his idea that we cannot know ourself. He uses the phenomenological reports from delusional patients to undermine the reports of non-delusional people to show that we are all deluded about our sense of self and our phenomenal world. For instance, Metzinger discounts phenomenal self-reports of sensory feelings of embodiment, claiming these are (must be) illusory because direct perception does not comply with a representational model of perception. At the same time, he uses the phenomenal self-reports of Cotard sufferers as evidence that embodiment is not an essential feature of conscious experience. He uses the reports of schizophrenic sufferers and those claiming to suffer from DID/MPD to demonstrate that the unitary self is an illusion created by the self model. While these conclusions may reflect the reality of their personal experience, they do not reflect the reality of the majority of personal experiences. It is inconsistent and poor research practice to dismiss self knowledge claims on one hand as being the product of an illusion, yet allow them on the other hand to support the claim that all is illusion.

As I discussed in Chapter 4, where I dealt with similar delusions of the self, there are limitations about what one can draw from the self-reports. The phenomenology does not map neatly onto specific neuronal nets to enable us to predict what phenomenological effects a certain type of damage will have. So while the delusions tell us that something is wrong, it won't tell us why it manifests the way it does. It won't tell us about the phenomenology. One also has to be cautious about taking the highly delusional phenomenological claims literally, especially when starting from a sceptical framework like Metzinger is. Yet this is precisely what he does. He says we should listen carefully to their accounts and take their phenomenology seriously because he thinks they will tell us something meaningful about what is going on in the 'normal' brain in relation to selfhood.<sup>109</sup> While these reports are a good indication that there are underlying problems that are having an impact on their sense

---

<sup>108</sup> Jaegwon Kim, *Physicalism, or something near enough*, Princeton University Press, Princeton, 2005, p. 4.

<sup>109</sup> Metzinger, 2003, p. 446/455. Zahavi expresses caution here and feels Metzinger takes the reports literally. Zahavi, 2005.

of self, they may not tell us much more. As I have said in earlier chapters, all that this evidence seems to indicate is that the self is vulnerable, that it can be damaged or even destroyed. Metzinger gives such reports more weight than is warranted because of his sceptical stance towards the numerous counter-claims of non-delusional subjects and to which he gives little credence.

I have argued that Metzinger cannot just dismiss our phenomenology as illusion, nor eliminate the conscious self the same way. While his PSM could do some of the self-work, by storing self-data, he doesn't provide a viable account of the initial identification of that data as belonging to self. At the same time, belonging is not the same as a conscious experience of. Metzinger dismisses the 'folk' self on the traditional grounds of lack of identity and singularity. He cannot find 'a self,' only a self-process. I am not convinced that these are enough grounds to dismiss the self. There is something process-like about the self, purely because the self and its identity are not 'given' at birth but develop over time. It may be that the self could be considered both an entity and a process (like a wave and a particle). According to Rescher, all objects can be considered both processes and objects simultaneously.<sup>110</sup> He discriminates between productive processes and transformative processes – those that produce an actual product or substance and those that transform the current state of affairs into an alternative state.<sup>111</sup> Selves could fit either of these definitions. At the same time, organisms are constituted by numerous processes; they grow and change and metabolise. As part of an organism, it is likely that the self will not have the same kind of existence criteria as other, more static objects. At the same time, though, the self should not be treated as exceptional in terms of normal epistemological constraints. We need to be cautious about placing more stringent requirements for identity and existence on selves than we expect of other objects in the world, a point I raised in Chapters 1 and 2.

In the next chapter, I address an alternative conception of the self as espoused by Galen Strawson. Like Dennett and Metzinger, Strawson is also a physicalist. However, Strawson argues that the self exists. In fact, he argues that the self can be considered a real concrete object with more ontological credentials than many

---

<sup>110</sup> Nicholas Rescher, *Process Philosophy: a survey of basic issues*, University of Pittsburgh, Digital Research Library, 2009, <http://digital.library.pitt.edu/cgi-bin/t/text/text-idx>.

<sup>111</sup> Rescher, 2009, p.28.

everyday items. This makes his position an interesting counterfoil to Dennett and Metzinger. Nevertheless, the self he posits is very minimal. I argue that it is too minimal to do the self-work it needs to do to be a viable candidate. My final chapter will discuss Damasio's self, which is, I claim, a compromise between the two fields. It offers a viable physicalist account of the self and its phenomenology without eliminating it.

### 5.13 Physicalism and reduction

Before I address Strawson's position, I want to briefly look at whether physicalism automatically entails reduction and, if it does, whether or not this entails the ontological negation of that reduced item.<sup>112</sup> Both Dennett and Metzinger posit physicalist accounts of the self and consciousness that seem to inevitably lead to an elimination of the very phenomenology they are trying to explain. I want to see if this is inevitable or whether one can hold a physicalist position that can retain things like mental states or consciousness.

Cognitive science has long held the view that we are our brains and that whatever we experience must, at some point, be explicable at least partly by what is going on in the brain. Physicalism is widely accepted as the standard model or means of explanation of all our human capacities, although there are those who are sceptical of this approach.<sup>113</sup> According to Beckermann et al, physicalism is the claim that 'there is nothing but physical objects and events' and that everything meaningful that can be said about these objects and events can be expressed in 'physical language.'<sup>114</sup> However, this seems to lead inevitably to an elimination of *mental* events, states and properties. Beckermann et al suggest that if this is to be avoided and physicalism retained, some kind of relation must obtain between the mental and the physical such that the one can explain the other. This relation is generally considered to be one of

---

<sup>112</sup> Once again, this is a big area to cover in any depth here. However, I think it is necessary to outline the relationship between explanation and existence as many of the discussions about the self claim that a physicalist explanation makes the self either eliminable or reducible to some other thing, hence not real or non-existent in its own right.

<sup>113</sup> In 'The mental problems of the many,' Unger argues against a form of 'scientificism' (a play on physicalism) that has as its metaphysical basis the idea that we are a 'highly complex wholly physical thing, with each of our powers just some sort of (physically derivative) physical power; or... each of us is epiphenomenal on, or supervenient on, a highly complex wholly physical thing. *Oxford Studies in Metaphysics: volume 1*, Clarendon Press, Oxford, 2004, p. 196.

<sup>114</sup> Ansgar Beckermann, Hans Flohr and Jaegwon Kim (eds), *Emergence or reduction? Essays on the Prospects of Nonreductive Physicalism*, Berlin, De Gruyter, 1992, p.1.

reduction.<sup>115</sup> While mental events can still be said to exist, they are ultimately explicable by recourse to more fundamental physical states and processes. The identity theory is one example of a 1:1 type mental-physical reduction. So physicalism does entail some sort of reduction of mental states to physical processes or states, but this reduction need not automatically entail eliminativism.

It would seem then that if there were an underlying physicalist explanation for the self and its phenomenology, as I suggested in the preceding chapter, then the self (and its phenomenology) could be described as that physical process or mechanism; even though it would be *reducible* to the underlying physical structures, without remainder. According to Beckermann et al, this should not negate its existence. It should still count as a self. In this thesis to date, I have referred to philosophers like Parfit, Carruthers, Dennett and Metzinger, all of whom have adverted to physicalist explanations of our self and its phenomenology. They have all concluded that this process eliminates the existence of the self and, in some cases, phenomenal consciousness as well. Two others that I discuss in the ensuing chapters, Strawson and Damasio, make the opposing claim.<sup>116</sup>

According to Churchland, reduction is ‘first and foremost’ a reduction between theories.<sup>117</sup> This means that if mental states are said to be reducible to brain states, then a theory of mental states (such as folk theory or psychology) should be reducible to ‘a theory describing how neuronal assemblies work.’<sup>118</sup> She goes on to say that such reductions not only produce ‘explanatory unification’ but lead to ‘ontological simplification’ as well. This is where the issue of reduction becomes relevant to the self, its potential existence and its ontological status. She claims that ontological simplification means that an item that is subject to an intertheoretical reduction may be considered to be the same as the item it is reduced to *or* may even be eliminated. Churchland cites ‘light’ as an example of simplification and ‘phlogiston’ as an example of elimination. It is still meaningful to talk about light but

---

<sup>115</sup> Beckermann et al, 1992, p.2.

<sup>116</sup> I discuss their positions in detail in Chapter 6 and 7.

<sup>117</sup> Patricia Churchland, *Neurophilosophy: towards a unified science of mind-brain*, MIT Press, Mass., 1986, p.278.

<sup>118</sup> Churchland, 1986, p.279.



ontologically it reduces to electromagnetic radiation.<sup>119</sup> Phlogiston, on the other hand, is not the same as oxygen and can be eliminated as a non-existent fantasy. Carruthers claims that the self does not exist because it is reducible to the actions of more than one mechanism or process. If we accept Churchland's description of reduction, this doesn't seem a legitimate position (elimination). Like water, the self might lose its ontological status but not its existence as an entity.

In a similar vein, Metzinger provides a physicalist explanation for the phenomenal self but claims that that self is an illusion. Both Dennett and Parfit claim that selves are *eliminated* by a reduction of this kind, rather than there being a theoretic simplification, although for different reasons. As discussed in Chapter 2, Parfit uses reduction to eliminate entities like selves and persons. They become purely linguistic entities or logical posits. For Parfit, the self is a fiction that has no physical underpinnings. Behaviour can be explained by recourse to mental events, without positing a self or person as well. There are just brains and bodies. On the other hand, Dennett claims his account is non-reductive. In *The Intentional Stance*, Dennett argues that beliefs and desires are not reducible to underlying causal neuronal structures, because there will be no 1:1 mapping of beliefs/desires to corresponding structures. If anything underlies these postulates, it will bear little resemblance to them. He claims that propositional attitudes do not exist as discrete causal entities and that folk theory is just false in this regard. The same is true of the self which, in line with Parfit, becomes a posit. Neither accounts are reductive in the sense outlined by Churchland, even though Parfit claims his is a reductive thesis. In both instances, the object being reduced has no corresponding match at the physical level of description. It doesn't appear to be an intertheoretic reduction. Both Parfit and Dennett treat selves like phlogiston, although Dennett claims that the self exists as an abstraction. It seems that the possibility of a physicalist reduction of the self is uniformly taken to demonstrate that it is neither real (no ontological status) nor existent in any concrete sense.

There are other physicalist positions that do not entail a strict identity theory of mental states or a straight reduction. Functionalism, for example, accepts that there may be different physical instantiations of the same mental property or state.

---

<sup>119</sup>Churchland, 1986, p.281.

Functionalism is the thesis that ‘mental states are defined in terms of their abstract causal roles within the wider information-processing system,’ not by their physical attributes.<sup>120</sup> This means that  $x$  may be considered functionally identical to  $y$  if it performs the same role, even if it is instantiated in a different form. While there will be some specific physical underpinning or realisation of a particular mental state leading to token/token reduction, there won’t be a type/type reduction. Both Dennett and Metzinger fit this model, claiming to be teleo-functionalists. According to Block, some functionalists argue that physicalism is wrong because there can be no type/type reduction.<sup>121</sup> This is because ‘there may be nothing physical that all pains share, nothing physical that makes a pain a pain.’ Churchland says that functionalists are often *antireductionist* because of this feature of multiple instantiability. In essence, functionalism can ignore the physical instantiation of states and focus instead on functional organisation and relations between inputs and outputs.

There are two points here that are relevant in discussing human phenomenology – epiphenomenalism and the explanatory gap. Functionalist accounts are generally medium-independent. They claim that intelligent robots could be constructed such that, in all relevant respects, they have similar mental lives to our own, even if the underlying hardware is radically different. Moderate successes in A.I. and robotics indicate that some cognitive activities can be replicated in other physical mediums. This adds weight to the idea that human cognition results from a certain kind of functional organisation of the neurological architecture instantiated in the brain. As was discussed in the previous chapter, parts of the brain do appear designed to ‘take care of’ specific information-processing tasks whose correct functioning seems to play a role in the kind of phenomenology we experience.

However, while computers and robots act in ways that seem to mirror, albeit in a limited way, our own capabilities, it is by no means apparent that they are conscious or that they are self-aware or that they ever could be.<sup>122</sup> Kierkegaard believed that investigating the self was outside the realm of the sciences. He believed that there

---

<sup>120</sup> Churchland, 1986, p. 351.

<sup>121</sup> Block, 2008, p. 4.

<sup>122</sup> There are two separate issues here, physicalism and the possibility of A.I. Some argue that only *living* systems are capable of consciousness. But even if robots can never be conscious, this does not preclude physicalist explanations of human consciousness.

was an explanatory gap between what science could say about the self and what the experience of subjectivity was like. Nagel is often taken as supporting a similar position.<sup>123</sup> The experience of the self is by nature subjective, as is the first-person perspective. Hence, the question of ‘what it is like’ to be a cognising, conscious human being does not seem to be fully captured by examining the activity of specific neural networks, just as talk of ‘C fibres firing’ or having a disposition to scream do not seem to capture what it is like to be in pain.<sup>124</sup> The problem for experiential states like pain is that neither the physicalist nor functionalist explanations appear to capture the felt qualities of pain in all its varieties and intensities. If we claim that such qualities are epiphenomenal, that they are produced by the underlying causal features, then such qualities become superfluous to explanation and have no causal role within the system itself. This apparent gap has been taken to mean either that science is not the correct tool for answering questions about what the self is or that qualia don’t exist or are non-causal.<sup>125</sup>

As such, functionalism appears to be a form of behaviourism which does little to explain the underlying causal mechanisms or the way things feel or seem, the phenomenology at the heart of the mental. Maxwell raises the explanatory gap argument against functionalism, while discussing its capacity to address the phenomenon of consciousness.<sup>126</sup> He claims consciousness will always sit outside of scientific explanation. His solution, however, is just as unsatisfactory as it places consciousness and qualia as extraneous to explanation. As Block points out, if mental states are treated as second order properties, then it is hard to see how they can ‘be causal or explanatory in a way appropriate to the mental.’<sup>127</sup> Kim also bemoans the

---

<sup>123</sup> See T. Nagel, ‘What is it like to be a bat?’ in *Philosophical Review*, 1974, pp. 435-50.

<sup>124</sup> See P. Churchland for an eliminativist account of qualia, *Matter and Consciousness*, MIT Press, Mass., 1988. Both Dennett and Metzinger dismiss the existence of qualia.

<sup>125</sup> See D. Chalmers for a discussion of the explanatory gap and the difficulties of resolving it. ‘Phenomenal Concepts and the Explanatory Gap’ in T. Alter and S. Walter. (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford University Press, 2006.

<sup>126</sup> Maxwell, 2011.

<sup>127</sup> Block, 2008, p. 8.

relegation of consciousness and qualia to ‘secondary qualities’ that either exist in some realm between the real and the unreal or are jettisoned as delusions.<sup>128</sup>

While treating the self as something non-physical or epiphenomenal places it beyond the scope of the hard sciences such as physics and neurophysiology, it is not an ideal solution as it also makes the self and its phenomenology non-causal.<sup>129</sup> I argued against one form of epiphenomenal self, the narrative self, in Chapter 3 for similar reasons. Yet in the preceding chapter I also argued that our self-phenomenology does appear to be subserved by physical structures in the brain, which would entail a reduction of some sort. Jaegwon Kim claims that simple physicalism always entails that mental states qua mental states are non-causal, whether via a process of reduction, eliminativism or epiphenomenalism.<sup>130</sup> He claims that a causally closed physical system appears to rule out mental causation in the physical world. Even a supervenience thesis has to entail reduction if it wants causation. Chalmers, for example, holds that phenomenal experiences are real distinct phenomena that are non-reducible to underlying causal mechanisms.<sup>131</sup> His position is a form of ‘naturalistic dualism,’ wherein mental properties can be said to be supervenient on the physical. He defines supervenience as follows: properties of type A are supervenient on properties of type B if and only if two objects cannot differ with respect to their A properties without also differing in respect to their B properties. Chalmers claims that mental properties supervene on physical properties (brain states) without being reducible to those physical properties (brain states) because the relation is not one of strict identity *or* functional identity.<sup>132</sup> This is because there are no physical/mental laws governing a symmetrical reductive relationship. According to Kim, supervenience and emergentism are seen as ways of retaining the autonomy of the mental without being antiphysicalist. However, Kim argues that if you want to retain mental causation, they still entail reduction. As he says, ‘the following principle seems highly plausible: *In order to cause a supervenient property to be*

---

<sup>128</sup> Jaegwon Kim, *Physicalism, or something near enough*, Princeton University Press, Princeton, 2005, p. 12.

<sup>129</sup> The felt quality of experience may serve an important function in alerting us quickly to the pleasantness/ unpleasantness of a stimulus, or as a warning that something is wrong.

<sup>130</sup> Kim, 2005, p. 11.

<sup>131</sup> See David Chalmers, ‘Naturalistic dualism’ in M. Velmans and S. Schneider (eds) *The Blackwell Companion to Consciousness*, Oxford: Blackwell, 2007, pp 359-368.

<sup>132</sup> D. Chalmers, ‘Naturalistic dualism,’ 2007.

*instantiated, you must cause one of its base properties to be instantiated.*'<sup>133</sup> While it might seem that 'M causes M\* to instantiate by causing P\* to instantiate', in actual fact P (on which M supervenes) caused P.\* So you either have two causal explanations or one is superfluous. If P (representing the physical) is the right one then, contra Chalmers, we have a reduction.

At the same time, Kim argues that we need to retain mental causation for two reasons.

First...the possibility of human agency... requires that our mental states have causal effects in the physical world. Second...the possibility of human knowledge presupposes the reality of mental causation.<sup>134</sup>

So while he acknowledges that mental causation is hard to retain within the physicalist/reductionist framework, it cannot be ignored; neither can mental states be explained away or eliminated. Churchland believes that mental states, propositional attitudes and consciousness are all potentially explicable by neuroscience. Against the antireductionists, she claims that reduction can be 'domain specific' which would negate the implications of multiple realisation.<sup>135</sup> If physicalism, and hence some form of reductionism, are true and consciousness and mental causation also exist, then the conceptual frameworks we have to date do not appear to be adequate to the explanatory task. While functionalism may be a non-chauvinist position, it may miss something critical. The medium may matter. Phenomenal consciousness may be a fundamental feature of living biological organisms. It may just be the way such organisms receive information.

It seems that physicalism does entail reduction but that reduction need not necessarily entail elimination. According to Churchland's model, though, it does seem to entail *ontological* elimination through simplification. Nevertheless, as demonstrated above, there is little consistency in how reduction is applied and in the conclusions drawn from that reductive process. Dennett claims his is a *non-reductive* position because the self (and its phenomenology) cannot be reduced to or explained by underlying mechanisms or by adverting to biology. His self is a socio-cultural

---

<sup>133</sup> Kim, 2005, p. 15.

<sup>134</sup> Kim, 2005, p. 10.

<sup>135</sup> Churchland, 1988, p. 357.

construct that overlays the underlying structures. Metzinger, on the other hand, uses the identification of underlying physical mechanisms to dismiss the existence of the self altogether, as does Carruthers.<sup>136</sup> For Metzinger, the discovery of mechanisms responsible for self-like activities does not entail claims that these constitute a self. Strawson would argue that this is a false move.<sup>137</sup> He claims that the metaphysical question of the reality of the self is constrained by human phenomenology. The phenomenology or sense of self sets out the necessary and sufficient conditions for something to count as a self. Strawson says that a self can be said to exist if something like the sense of self is an accurate representation of it. So whatever that representation is true of, that thing counts as a self. Conversely, if there is such a thing as a self, then some sense of self will be an accurate representation of it. This move disallows arguments to the effect that one could claim that there is a self but the sense of self does not accurately represent it, or that the sense of self accurately represents something but that something is not a self. As is apparent in the previous discussion, both Metzinger and Dennett make this move.

My own view is that the self is a complex and is not reducible to a single mechanism or even a set of mechanisms. This may make it an emergent entity which is still explicable by recourse to underlying physical structures but which cannot be reduced to their actions in isolation from each other. This may be in much the same way that liquidity is a property of water but is not a property of hydrogen or oxygen, while water is still reducible to H<sub>2</sub>O. Given what we know so far, the self is likely to be a system or a network of systems, rather than a singular neuron or even a singular process. I will discuss this possibility in the final chapter.

---

<sup>136</sup> G. Carruthers, 'A Model of the Synchronic Self,' *Consciousness and Cognition*, Vol 16, 2007, pp. 533-550.

<sup>137</sup> See Galen Strawson, 'The Self' and 'The Self and the SESMET' in *Models of the self*, eds. S. Gallagher and J. Shear, Imprint Academic, UK, 1999, pp 1-24; and Galen Strawson, 'The Self' in *Journal of Consciousness Studies*, Vol. 4, nos 5/6, 1997.

## CHAPTER 6

# STRAWSON'S SELVES

What was I before I came to self-consciousness?... I did not exist at all, for I was not an I. The I exists only insofar as it is conscious of itself... The self posits itself, and by virtue of this mere self-assertion it exists.<sup>1</sup>

### 6.1 Introduction: an alternative position

In the last chapter, I looked at two physicalist accounts of the self that argued that the self was variously an abstraction, an illusion, a fiction or a simulation. Both argued that the Cartesian self was non-existent and that the self we think we have is a late addition to or product of our particular cognitive architecture. I argued that neither account was viable as an explanation of our self-phenomenology, and that their respective models could only work if they demonstrated that phenomenal consciousness was an illusion. In addition, both Dennett and Metzinger hold ontological positions such that they disallow any physical mechanism or complex of mechanisms as possible candidates for a self. This means that there can be no alternative to a Cartesian self or that what is called by the name 'self,' is either epiphenomenal, illusory or reducible to something that is a non-self.

In this chapter, I want to address another physicalist conception of the self, that posited by Galen Strawson.<sup>2</sup> The difference between Strawson's account and those of Dennett and Metzinger is that he appears to be positing a viable alternative to the Cartesian self. Like Dennett and Metzinger, Strawson is overtly physicalist. However, contrary to their claims, he believes that the self is both real and concrete and that it has a locatable physical manifestation. He not only retains many of the Cartesian self-properties, such as mentality and singularity, but insists that our self-phenomenology be used both as a constraint and a litmus test for any theory of the self. As such, his account is an interesting foil to Dennett's and Metzinger's.

---

<sup>1</sup> J. G. Fichte, *The Science of Knowledge*, trans. Peter Heath and John Lachs, CUP, 1982, p.1794-5.

<sup>2</sup> Strawson has written a series of articles on the existence of the self, culminating in a comprehensive explication and defence in his book, *Selves: an essay in revisionary metaphysics*, OUP, 2009. In this tome, there are some notable refinements and slight deviations from his earlier accounts which are important to note. These changes mean that his final overall conclusion resembles my own thesis much more than his earlier works did. See also G. Strawson, 'The phenomenology and ontology of the Self', *Exploring the self*, D. Zahavi (ed), John Benjamins Publishing Company, Amsterdam, 2000; 'The Self' and 'The Self and the SESMET', *Models of the self*, eds. S Gallagher and J Shear, Imprint Academic, UK, 1999, pp 1-24 and pp 483-518 respectively; and 'The Self', *Journal of Consciousness Studies*, 1997, 4, No. 5/6, pp.405-428.

This chapter is a close analysis of Strawson's position, his approach and his rationale. There are some similarities between his position and my own. He, like me, is trying to establish what the term 'self' entails; whether or not such things as selves exist; and what the metaphysical status of such things would be if they did. He, like me, is unequivocal in his response; selves exist. He, like me, claims that selves can be considered real existent concrete entities. I have drawn on parts of his methodological framework and phenomenological constraints throughout this thesis and in the preceding chapter to demonstrate how Dennett and Metzinger go outside those constraints to argue against the self.

However, there are strong points of disagreement between my own position and Strawson's. While I am sympathetic to Strawson's intention, I do not support his main contention that the self is non-extended in time, nor do I support his view that our sense of self is completely disembodied or purely mental. The self Strawson defends ends up being a (very brief) moment of subjectivity akin to Fichte's and Nozick's emergence of the 'I' in that moment of self reflection. I conclude that Strawson's self is so minimal it is unlikely to be able to support the robustness of our self-phenomenology. As such, one might claim it counts as a moment of subjectivity but it is too fleeting to count as a self, no matter how minimal.

## **6.2 Strawson's position**

In his earlier works, Strawson has consistently argued that the self is real and that it can be considered an object in its own right. In *Selves*, he puts forward the stronger case that selves are 'concrete, physical things of a certain specific sort, and ... they qualify as objects.'<sup>3</sup> The self that Strawson defends is a 'minimal self,' what he calls a SESMET or 'thin subject.' This is a momentary conscious self, consisting of an experience of oneself as a single mental subject of experience. He argues that such selves can be considered real, concrete, physical objects much like any other existent physical object.

Strawson's selves are not like the every day 'folk' conception of the self, assuming that there is one. In that sense it is a counter-intuitive model. A Strawsonian self is very minimal, lasting at most 3 seconds and possibly as little as 0.3 seconds. In

---

<sup>3</sup> Strawson, 2009, p. 13.



addition, there are multiple selves of this kind occurring serially in quick succession in any one conscious human brain, each one discrete with its own identity conditions. According to Strawson, this rather unusual claim is supported by reflection on one's own phenomenology. He claims we don't directly experience a continuous self but rather infer it from the continuity of our bodily existence. What we experience directly are moments of subjectivity or 'I' moments.

... when I consider myself in the whole-human-being way I fully endorse the conventional view that there is in my case – that I am – a single subject of experience – a person – with long-term diachronic continuity. But when I experience myself as an inner mental subject and consider the detailed character of conscious experience, my feeling is that I am – that the thing that I most essentially am is – continually completely new.<sup>4</sup>

In some of his earlier work he has described this as the 'pearl necklace' view of the self, although he does not use this analogy in his latest book.<sup>5</sup> A Strawsonian self has no long-term continuity or diachronic identity. It is a purely synchronic self (although it is minimally extended). A new self is created at each moment of a SESMET (sometimes called a chronon to indicate a small measure of time) which is an experience of oneself as an experiencing subject. According to Strawson, this minimal self is, nevertheless, a real concrete object.

Both of these claims are contentious. While I agree that the self can be considered a physical object in much the same way that Strawson does, and that his minimal self could count as such an object, I do not support the idea of a minimal transient self. As will be demonstrated in the following sections, Strawson's self is too discrete, too minimal and too multiple. Nevertheless, he directly addresses and incorporates elements of our self-phenomenology in ways that Dennett and Metzinger do not. He also provides a grounding for discriminating between subjectivity and selfhood that I draw on in the final chapter of this thesis. It may well be that there is a spectrum of subjectivity from minimal to maximal, with human selfhood as a maximal form of self-conscious subjectivity. The minimal subject, though, should not be confused with Strawson's minimal self. As alluded to in earlier chapters of this thesis, while several writers use the term 'minimal self,' it does not always describe or capture the same thing. Strawson's minimal self is not the grounding on which a maximal self

---

<sup>4</sup> Strawson, 2009, p. 247.

<sup>5</sup> See G. Strawson, 'The phenomenology and ontology of the Self', *Exploring the self*, 2000.

can develop. It is not a basic biological or core self that leads to a psychological self narrative. Strawson's minimal self is the minimum required for a sense of self to still exist once inessential features of our self-concept have been stripped away.

### 6.3 Strawson's method

Strawson starts his analysis with the phenomenology of the self. According to Strawson, one's metaphysics of the self should be constrained by the phenomenology of the self, as it is the phenomenology that is the root of the problem, what creates the problem of the self that needs to be explained. So his starting point is an explication of the sense of self, what in his earlier work he called the sense of a mental self (SMS).<sup>6</sup> He begins with a group of properties that he claims capture the more or less universal sense of a self. These are a sense of the self as 1) a thing, 2) a single, mental thing, 3) synchronic, 4) diachronic, 5) ontically distinct, 6) a subject of experience, 7) an agent, and 8) having a personality.<sup>7</sup> While he acknowledges that selves may well have non-mental properties as well, he argues that these are largely irrelevant to one's *experience* of oneself as a self. This experience is typified by being a mental self that is not wholly (or even partly) considered identical to one's body, which is viewed as physical.

Strawson spends some time drawing a line between the mental and the non-mental to establish that our sense of a self is very much of a mental self. This is partly to define or focus on what he believes is the pure sense of a self and partly to counter those philosophers who push for an embodied or purely physical existence. That position, he suggests, denies the reality of lived experience as a conscious functioning human being. Self-reflection does appear to lend some support to Strawson in this. When we are most conscious of our self, there is always an inner mental component.

Being self-aware seems to entail awareness of self as a mental entity in that it involves thought. Thoughts can come into our head unbidden, momentarily demanding our attention away from our current physical activity. This mental state is in contrast to those states of awareness when we are aware of external stimuli or events that draw our attention away from ourselves. In that state, we can be so

---

<sup>6</sup> see Strawson, 1999, pp 1-24 and pp 483-518 respectively.

<sup>7</sup> Strawson, 1999, p. 3.

engrossed that we lose self-awareness. This does not fit Strawson's criterion for selfhood. The self exists in self-reflection, thus mentality. Strawson does include emotions and moods as part of that mental life, even when these are generated by bodily conditions.<sup>8</sup> Despite this, Strawson claims we are more mental than anything else. He claims that everyday experience is of an 'extraordinarily rich, rapid, nuanced, complexly inflected, interdipping flow...[of] inner mental goings-on *experienced as such*' (his italics).<sup>9</sup>

At the same time, Strawson is sceptical of the division of the world into the opposing categories of mental and physical properties. If one is a materialist, and Strawson places himself firmly in this camp, then all properties, objects, events and processes are ultimately physical. The only real distinction (and he means this ontologically) is between the mental and the non-mental which he classifies as the distinction between the experiential and the non-experiential.<sup>10</sup> An experience for Strawson is always a conscious, inner, hence mental, experience. While this position assumes that all experience is mental, it does not assume that all mental states are conscious. Strawson does contemplate being conscious of something (x) subliminally, which implies not being aware that one is conscious of x.<sup>11</sup> Nevertheless, he does claim that all, and only, experiences are what constitute the self. Wilkes, in particular, brings him to task over this claim which I address in a later section.<sup>12</sup> In addition, Strawson claims that *self* experiences are always mental, never physical. They are never experiences of the body. He does not support the idea of an embodied self, what he refers to as the EEE position (environmentally embedded, embodied, ecological, enactive).<sup>13</sup> I deal with this point in more detail later in the chapter.

Contrary to Dennett and Metzinger, Strawson accepts our phenomenology as real as opposed to illusory. According to Strawson, we really are conscious individuals and there are things that we directly experience, including the world of external objects.

---

<sup>8</sup> Strawson, 2009, p. 25.

<sup>9</sup> Strawson, 2009, p. 26.

<sup>10</sup> Strawson, 1999, p. 6.

<sup>11</sup> Strawson discusses this briefly in 'Real direct realism: reflections on perception' *The Nature of Phenomenal Qualities*, ed. P. Coates and S. Coleman, (OUP, forthcoming, 2014), p.3.

<sup>12</sup> K. Wilkes, 'Know Thyself,' *Models of the self*, eds. S. Gallagher and J. Shear, Imprint Academic, U.K., 1999.

<sup>13</sup> See in particular, Strawson, 2009, p.23

There are self-as-subject-experiences. He supports a form of direct realism, claiming that experiencing a conscious mental representation just is being in direct perceptual contact with an object.<sup>14</sup> As he says,

A perception of an object must involve a mental representation of the object, but the representation doesn't get in the way. On the contrary. When a subject perceives an object, the subject's having a mental representation of an object is an essential part of what constitutes the subject's direct perceptual relation to the object.<sup>15</sup>

Similarly, we are directly aware of our own inner states and thought processes. These all result in inner or mental experiences. While Strawson doesn't discount the bodily sensations which Damasio and others claim ground the sense of self, he does state that these mostly remain in the background and are less significant than mental experiences.<sup>16</sup> It doesn't define the self-experience. Ultimately, the self is experienced as a predominantly mental thing.<sup>17</sup>

Given that there really is a phenomenal sense of self, Strawson then claims that any theory about the self has to accommodate that phenomenology. More precisely, the phenomenology should constrain what the self can and can't be. He originally set up his phenomenological constraint in the following two-fold equivalence claim, which I referred to in Chapter 5:

(E1) If there is such a thing as the self, then some SMS (sense of a mental self) is an accurate representation of something that exists.

(E2) If some SMS is an accurate representation of something that exists, then there is such a thing as the self.<sup>18</sup>

This claim was to prevent philosophers from saying that there is a self but it bears no resemblance to our experience of it, or that there is a sense of self, but what it represents is not really a self.<sup>19</sup> What Strawson is saying is if one could show that a

---

<sup>14</sup> Strawson says that 'Mental representation is mental presentation is mental registering.' 2014,' p. 7.

<sup>15</sup> Strawson, 2014, p. 7.

<sup>16</sup> I have referred to this sense in previous chapters. See i.e. *The Inner Touch: archaeology of a sensation* by Daniel Heller-Roazen, Zone Books, New York, 2007 for a history of the bodily sense.

<sup>17</sup> I am not convinced by his dismissal of the bodily sense and address this in some detail in my own account of the self in this chapter and in Chapter 7.

<sup>18</sup> Strawson, 1999, p. 5.

<sup>19</sup> See Chapter 5 for a discussion of how Dennett does exactly this in *Consciousness Explained*, 1991, where he claims that our phenomenology is based on an illusion and that what subserves that illusion is nothing like a self. See Metzinger, for a similar move re the PSM.

sense of oneself accurately represents something in the cognitive system then that thing in the cognitive system would count as a self and vice versa.

In his latest work, he puts forward a slightly different version of the equivalence claim. To avoid confusion, I will represent the second set of equivalence claims using italics.

*(E1)* If there are such things as selves, then they must have the properties attributed to the putative self in any (every) genuine form of SELF-experience.

*(E2)* If there exist things that have the properties attributed to the putative self in any (every) genuine form of SELF-experience, then those things are selves.<sup>20</sup>

The main difference between the two sets of claims is the replacement of the word ‘representation’ with ‘properties.’ It seems as if the second set of equivalence claims takes a more direct realist position by removing the term ‘representation.’ This makes the experience of a self no longer a derived sense of a self, but rather a ‘presentation’ of a self that is directly experienced as having certain properties. It seems to make the claim *(E2)* easier to establish in that if one were to find some thing with those properties, then that thing would be a self almost by definition. Interestingly, Strawson also removes any reference to a sense of self in both *(E1)* and *(E2)*, instead using the term SELF-experience. This is a divergence from his earlier work which had the SMS (sense of a mental self) as a major component of his theory. In fact, in his earlier work Strawson was at pains to distinguish between the sense of self (what it is like to be me) and the sense of the self, (what is it like to experience something from a position of selfness).<sup>21</sup> This original distinction mirrored more closely my own separation of personal identity from selfness. The only reason I can think that he has removed any reference to a *sense* of self is that one’s sense of (a) self could contain individual variations or broader features than are contained within the epithet SELF-experience. For instance, a sense of (one’s) self could have additional features, such as diachronic continuity or personality or features outside the mentality constraint.

---

<sup>20</sup> Strawson, 2009, p. 56.

<sup>21</sup> See Galen Strawson, ‘The Self’, *Models of the Self*, 1999.

Often our sense of self is of a particular kind of self and does entail aspects of our personal identity. Strawson wants to focus on the bare bones of SELF-experience to find the essential self. He wants to reduce the self-phenomenon to its minimal features. A focus on SELF-experience enables Strawson to deal with what he believes is contained directly within that experience, nothing more. What changing E1 and E2 into *E1* and *E2* does, is allow him to claim that the self and self-experience are aspects of the same thing. They seem to collapse into one, much in the same way his object and its properties do.<sup>22</sup> The self is fully captured by the phenomenology, the SELF-experience.

I am not sure that this is paying tribute to the self phenomenology in the way Strawson claims he is. Self-traits are experienced subjectively, from the inside, and from what is often referred to as a privileged position. Our knowledge of ourselves is experiential; we are the subject of our experiences not the object. This has led to claims that, as an acting self, we don't really experience ourselves *as a self*. We don't experience or observe our *self* objectively, as an object. As I mentioned in Chapter 2, this view is called the 'elusiveness thesis.'<sup>23</sup> Cassam claims that any physicalist conception of the self necessitates awareness of self as *object*, not just as subject. If the self is not an object in the world then it doesn't exist.<sup>24</sup> Strawson argues that the self is an object but it is not clear we experience it as such. If it is an object, its objectness is likely inferred rather than experienced directly.

## 6.4 The SESMET

As in his earlier work, Strawson reduces his eight features of the self of experience, those comprising the total sense of a self, to just four essential features or properties, using what he calls his Whittling argument. Strawson is concerned with what we really experience as aspects of our self-experience. He also wants to find out what aspects of that experience are essential to having a self-experience and those that are not. Personality or personal identity are not essential given that, in cases of amnesia, sufferers can still experience subjectivity without knowing their identity. I made

---

<sup>22</sup> See his discussion of the inseparability of an object from its properties in *Selves*, p. 615.

<sup>23</sup> See Quassim Cassam, 'Introspection and bodily self ascription,' *The Body and the Self*, ed Bermudez, Marcel and Eilan, MIT Press, 1995.

<sup>24</sup> This is too strong as there are many things that can be said to exist that are not classified as objects. Here I believe he is adverting to Lowe's *concrete* object.

similar claims in Chapter 2. He also dismisses agency as an essential feature for self-experience. One can, he claims, separate oneself from one's body in one's thinking without loss of subjectivity. One can also experience subjectivity when there is no capacity for agency, as in cases of paralysis. He also argues that our belief in our agency exceeds what is, in fact, the case. Thoughts occur to us more than we intentionally create them.<sup>25</sup> Strawson concludes that the self-experience is predominantly mental, an inner experience that is distinct from the physical. This experience is of some *thing*, rather than of a property or a process. This thingness implies singularity. However, Strawson says that the singularity of self-experience is not just of a single unit or collection of things but of strong internal unity. It is more like a marble than a pile of marbles.<sup>26</sup> In addition, this self-experience is, by definition, subjective. The self-experience is as a subject of experience.<sup>27</sup> This gives us a single, mental subject-of-experience thing.

Strawson makes it clear that he is drawing a distinction between one's experience of oneself as a subject of experience and experiencing oneself as a whole human being. *Self-experience* is 'a certain way of experiencing oneself as a subject of experience' particularly insofar as one has 'mental being.'<sup>28</sup> He claims we use 'I' to refer to two distinct selves, oneself as the inmost subject of experience and oneself as a whole embodied human being. To reflect this, Strawson makes a distinction between the use of 'I,' 'myself' (etc.) and the use of 'I\*', 'myself\*' (etc.) to represent the differing uses of 'I'. In this way, the 'I\*' differentiates the experiential, mental I from the whole-bodied I of 'I went to the shop.' It is the inner I\* that is the self of experience. Accordingly, Strawson posits that self-experience is experiencing oneself as 'subject-of-experience-as-single-mental-thing' or SESMET.

To figure oneself as a SESMET is...to figure oneself as a single thing or object specifically when one is figuring oneself as a subject and figuring oneself figured as a subject specifically in respect of one's mental being.<sup>29</sup>

---

<sup>25</sup> Strawson, 2009, p. 189.

<sup>26</sup> Strawson, 2009, p. 72.

<sup>27</sup> I have also argued for this in Chapter 2. Strawson makes a similar point that experiences entail experiencers; see S Shoemaker, 'Parfit on Identity,' *Reading Parfit*, Jonathan Dancy (ed.), Blackwell Publishers, Oxford, 1997.

<sup>28</sup> Strawson, 2009, p. 205.

<sup>29</sup> Strawson, 2009, p. 206.

Strawson then deductively argues that if self-experience is equivalent to SESMET-experience, then self is equivalent to SESMET. This is an interesting move because it collapses the sense of a self or self-experience into the self. Rather than the self-experience implying that there is a self having or undergoing that experience, this move seems to imply that the experience (SESMET) *is* the self. It seems to be saying ‘I experience myself as a self and that is what a self is.’ In his defence, one could argue that what Strawson is doing is embedding *self-reflection* into the self-experience, such that one is experiencing oneself as an experiencing subject. It is this that collapses into the self, rather than just the experience. Strawson does draw parallels between his own account and that of Nozick’s so this may well be the case. I will expand on this point in more detail later on.

Strawson wants to limit the self\* of self-experience to those features which appear defensible *in experience*. He claims that this does not include the self\* as persisting in time. Thus, Strawson’s account of self-experience is of a singular moment of experience, what he calls the ‘lived present of experience.’<sup>30</sup> While such a moment is not durationless, its temporal extension can be limited to under a second. To be synchronically singular, the self\* has to be experienced as singular or unified within the duration of that lived present of experience, whatever its duration is. In earlier texts, Strawson referred to this as an ‘uninterrupted or hiatus-free period of consciousness.’<sup>31</sup> In contrast, diachronic singularity requires that the self\* be experienced as single or unified across (more than one) lived present of experience. Strawson claims this does not accord with reality or lived experience. While Strawson could accept a longer than 0.3 second self, he rejects the idea of a persisting self, one that has relatively long-term duration. This idea, he claims, is not experiential but inferred. Drawing on his own self-phenomenology, he claims he feels no sense of oneness with his past selves\* or possible future selves\*. As he says, ‘[m]y past is mine\* in the sense that it belongs to me\*, but I don’t feel I\* was there in the past.’<sup>32</sup>

---

<sup>30</sup> Strawson, 2009, p. 75.

<sup>31</sup> Strawson, 1999, p. 21.

<sup>32</sup> Strawson, 2009, p. 228.



A Strawsonian self\* only exists in the moment and at that moment of lived presence. Strawson's next move is to show that there really is a self or SESMET with the features he has identified as representative of that self.

## 6.5 Strawson's metaphysics

There are two major issues facing the supposed reality of the self. One is establishing that it exists and the other is establishing the nature of that existence. As discussed in Chapter 1, existence alone does not seem to lead to metaphysical reality; this will depend, in part, on the nature of that existence, on the kind of thing it is. If selves are considered to be entities like nations or events, then their existence might be considered reliant on a socio-linguistic convention. If they are considered an entity like colour or a posited entity like Dennett's centre of narrative gravity, they may be seen as abstract. Or they might have an obvious concrete reality in the way a human being or tree does, or even an atom. As I have discussed before, opinions vary as to the ontological status of even some ordinary everyday entities like apples or leaves. It seems that one's general ontological commitments tend to determine how one categorises the self. Similarly, Strawson's approach to the self reflects his own ontological commitments. He starts by analysing what is considered to be the basic unit of existence, the object. He claims that his *sysele* or 'thin subject' is as much an object as anything else, given that it is a natural 'indecomposable unity' physically instantiated in a locatable place in the brain (I discuss this point in detail in the next section). He appears ambivalent, however, about whether there are such things as objects in the world or whether everything can be considered a process.<sup>33</sup> He claims that there is 'no *metaphysically weighty* distinction between objects and processes,' and that all physical objects are dynamic entities constituted 'out of time-matter, process-stuff.'<sup>34</sup>

This is an interesting position and is reflective of an ontological position called Process Philosophy (spearheaded by Whitehead), that puts forward the idea that everything can be seen as a process of some kind. Rescher is the modern day

---

<sup>33</sup> See Strawson, 'What is the relation between an experience, the subject of the experience, and the content of the experience?' *Philosophical Issues*, 13, *Philosophy of Mind*, 2003, pp. 287-290, where he discusses his views on objects as processes in more detail.

<sup>34</sup> Strawson, 2003, p. 288.

advocate of process philosophy.<sup>35</sup> The idea is that processes need to be part of any ontology to explain change and becoming. Processes produce outcomes, many of which are substances or changes to substances. Natural processes are the paradigm case. According to Rescher, a natural process ‘inherently exhibits a structure of spatio-temporal continuity.’<sup>36</sup> He argues that by placing objects at the centre of one’s ontology, process becomes secondary or subordinate in importance. This has the effect of putting persons and agents into a secondary category to objects because they are not ‘substantial enough’. Thus, not only would something like the self be ontologically downgraded but so would processes per se. As was discussed in Chapter 4, when Metzinger and Dennett refer to the self as a process, they are, in fact denigrating processes to an abstract or non-existent category of things. Rescher goes on to argue that ontologies that focus on objects and properties (inert material) lead to a reality divided into discrete temporal stages because process has been removed.<sup>37</sup> Time becomes a location in space and change is static

Like Rescher, Strawson acknowledges the dynamism of nature and the fundamental role of processes. At the same time, he argues that *if* there are such things as physical objects, and these are one’s benchmark of real metaphysical existence, then his minimal self would qualify as such an object. In fact, he argues that his minimal self is a paradigm case of objecthood. It qualifies as an object when a human being does not.

Strawson argues that for something to count as a singular object, it must have strong internal unity. He discounts brain-system theories of the self precisely because they would lack the necessary strong unity that is required for singularity and objecthood. As someone supporting a system view of the self, I do not believe this is the case, as I indicated in Chapters 1 and 2. I address this point again in a later section of this chapter. Nevertheless, he does put forward a strong case to demonstrate that the single moment of subjective experience, the lived present of experience, is not decomposable. He claims it has the kind of object-unity prescribed by van

---

<sup>35</sup> Nicholas Rescher, *Process Philosophy: a survey of basic issues*, University of Pittsburgh, Digital Research Library, 2009, <http://digital.library.pitt.edu/cgi-bin/t/text/text-idx>.

<sup>36</sup> Rescher, 2009, p. 23. Interestingly, Rescher blames the elder Strawson for giving ontological priority to objects and ignoring processes.

<sup>37</sup> Rescher, 2009, p. 36.

Inwagen.<sup>38</sup> This is more unity and singularity than most of our everyday objects can lay claim to. This lived present will be realised in the brain in some way as a single phenomenon. This realisation is the I\*. It exists, but only during that moment of experience. Given the subjective nature of the I\* or self\*, it can only exist during moments of conscious experience of itself. These are very brief, rather than continuous. The I\* is created anew at each moment of lived present. Thus, in any given lifetime there will be multiple selves\*, even if there is only one I. Selves\* exist in addition to human beings. However, they have minimal persistence conditions; each self endures for only a fraction of a second. On the other hand, the I as the human being and as a subject of experience does persist (and has persistence conditions).

Strawson claims that a mental self (self\*) exists at any given moment of subjective experience. This self-moment is an experience of singular subjectivity. It is represented by a 'set of neuron-and-neurotransmitter-...constituting atoms or fundamental particles in a certain state of activation.'<sup>39</sup> Strawson is not particularly bothered about the exact details of this activation pattern. He is merely making the point that there will be (or so he surmises) some one physical instantiation of that experience somewhere in the brain that would constitute an object-like singularity. This does not dismiss the phenomenology by proffering a simplistic reduction but merely claims that there will be some physical state of activation that will subserve this experience of selfness. A concrete physical manifestation removes any grounds for classifying the self\* as an abstract object. He claims that it is 'as real as a rabbit, as much an object as a grain of salt and as physical as a jackhammer.'<sup>40</sup> Strawson even intimates that the minimal self-moment, the neuron cluster, may be able to exist independently of that human being or the whole brain, depending on how much of the human being can be progressively reduced (surgically or imaginatively) without impacting on that lived present of experience.

---

<sup>38</sup> See P. van Inwagen, 'Symposium on *Material Beings*,' *Philosophy and Phenomenological Research*, Vol. 53, pp. 683-719.

<sup>39</sup> Strawson, 1999, p. 21.

<sup>40</sup> Strawson, 1999, p. 21.

## 6.6 Strawson's argument for selves as syseles

It is probably important at this point to include a brief account of Strawson's argument that establishes the identity relationship between selves, SESMETs and syseles or thin subjects, as this is critical for Strawson's position that selves exist as objects. Strawson puts forward what he calls his *thin* conception of the subject (sysele). In its nuanced form, it is represented as the following.

a subject of experience is an inner thing of some sort that exists if and only if experience exists of which it is a subject.<sup>41</sup>

Strawson draws support for this position from Descartes, Kant and James, all of whom claim the same, albeit substituting conscious experience for thought or thinking.<sup>42</sup> He claims that there can't be a *subject* of experience unless 'some *experience* exists for it to be the subject *of* (his italics).'<sup>43</sup> Strawson likens the necessity of this condition to that of a box without a surface. The property and the object are existentially inseparable, even if they can be separated conceptually. If experiences exist, then thin subjects exist. According to Strawson, this entails that if there is no experience there is no thin subject. Given that experience is not a plenum, then the subject of experience will cease to exist when experience ceases.<sup>44</sup> In his review of Strawson's book, Shoemaker takes issue with this conclusion. In the past, Shoemaker has defended the claim that experiences entail experiencers, in much the same way that Strawson claims experiences of self entail a (thin) subject. In his critique, Shoemaker argues that Strawson doesn't demonstrate that the reverse is true, that the thin subject can't or doesn't exist without an experience.<sup>45</sup> Shoemaker is not sympathetic to Strawson's main claims about the existence of selves, but his point here is a relevant one. Strawson relies on immediacy and the lack of continuity of the self moment. As I mentioned earlier, he collapses the experience into the experiencer as the one thing or object. For Strawson, the existence of the thin subject

---

<sup>41</sup> Strawson, 2009, p. 324.

<sup>42</sup> See W. James: 'My present Thought stands thus in the plenitude of ownership of the train of my past selves, is owner not only *de facto*, but *de jure*, the most real owner there can be, and all without the supposition of any 'inexplicable tie,' but in a perfectly verifiable and phenomenal way.' The Principles of Psychology (1890), p. 361, *Classics in the History of Psychology*, accessed April 2011, <http://psychclassics.yorku.ca/James/Principles/prin10.htm>.

<sup>43</sup> Strawson, 2009, p. 325.

<sup>44</sup> Interestingly, Dennett talks about the 'gappiness' of consciousness to argue that we are not phenomenally conscious.

<sup>45</sup> Sydney Shoemaker's review of Strawson's *Selves*, [Notre Dame Philosophical Reviews](#), 2013.

(sysele) is reliant on the existence of a self-experience. The self-experience (property) is realised by the sysele (object). Yet selves may well be underpinned by other factors such as memories, bodily sensations or patterns that maintain a sense of self without the constant experience of some thing. As Shoemaker claims, that could be the subject him/herself.

Accepting that Strawson has established that the thin subject exists, he then sets out to show that it is 1) an object, to give it metaphysical existence and 2) it is a SESMET, to link it to the self. The moment of experience is non-decomposable and physically instantiated, thus giving it object status as singular, unified and concrete. He then argues that syseles have all the properties associated with SESMETs. By arguing forward he can then establish that syseles are SESMETs and SESMETs are selves. If syseles exist, then selves exist, as illustrated in the diagram below.

If Syseles  $\longrightarrow$  (if) SESMETs  $\longrightarrow$  selves

## 6.7 Critiquing Strawson's minimal self: points of agreement

I support Strawson's methodology and have drawn on it to further my own argument. Discussion of selves must start with the phenomenology and be directed by that phenomenology. This is at the heart of what it is to experience a sense of self. It is too easy to dismiss the problem of selves by discounting that phenomenology as illusory. Arguing that it is illusory means one does not have to be constrained by it or take it into account when proffering theories of the mind or consciousness. This avoids the problem rather than solving it.

It doesn't explain why we experience the phenomenology we do, even if it is illusory, nor why it is remarkably uniform across individuals from differing socio-cultural backgrounds. Both Dennett and Metzinger deny the phenomenology; it only seems as if one is conscious or having experiences; it only seems as if one is self-conscious. For Dennett, this is the user-illusion of the meme machine. Metzinger argues that our experience of the world is indirect and, as a consequence, is an inference or judgement. He claims we are mistaken about the nature of our experience and, hence, the subsequent belief in a self or an experiencing subject is

false.<sup>46</sup> The phenomenal self is an illusion. However, Metzinger doesn't deny that there are experiences, and this fact entails, using Strawson's constraint, that there are subjects of experience. Strawson accepts that we do have experiences and that they are the way we describe them as being. For his purposes he does not need them to be veridical accounts of the external world or to represent accurate self-knowledge. As I argued in Chapter 5, the truth or falsehood of the content of experience does not negate the experience of oneself as the subject of that experience. The self could still exist even if, to follow Descartes' meditation, one were continually fooled by an evil demon.

There are, however, problems inherent in using phenomenological accounts in that they are, by definition, subjective. This means they are difficult to independently verify and to compare or contrast with each other. There is also empirical evidence to show that they can be unreliable, as Dennett points out.<sup>47</sup> While it does appear as though all living humans are conscious and have phenomenal experiences, and that these experiences are from a singular point of view, it is by no means obvious what other commonalities there are. This problem is brought into focus when Strawson describes his own self-phenomena. He claims to experience no sense of himself\* as the same self across those singular moments of lived experience. He feels no association with his past selves\* even though he\* owns the content of their thoughts. He claims that any continuity he has is not experiential but derived. It belongs to himself as a human being not himself\* as a self. It would appear from this that diachronic singularity or unity is something that can be experienced differently by humans, with something like Hopkins' strong unity at one end of the spectrum and Strawson's momentary self\* at the other.<sup>48</sup> It is not clear how or if this can be resolved. Having said that, the four basic characteristics of subjective experience suggested by Strawson do appear sufficient to warrant a minimal universal self-experience. In fact, many neuropathologies such as dementia or Alzheimer's can

---

<sup>46</sup> T. Metzinger, *Being no-one: The self model theory of subjectivity*, MIT Press, Mass, 2003.

<sup>47</sup> See Dennett for a detailed account of the mistakes we can make in 'Qualia Disqualified,' *Consciousness Explained*, 1991.

<sup>48</sup> I use the same quote from Hopkins to open Chapter 7 that Strawson used to open his paper on selves. It captures the poignancy of Hopkin's phenomenal experience of self. '...Nothing in nature comes near this unspeakable stress of pitch, distinctiveness, and selving, this self-being of my own.' Gerard Manley Hopkins, *The Poems of Gerard Manley Hopkins*. Ed. Robert Bridges. London: OUP, 1930, p.123.

deplete one's self capacities without necessarily destroying one's subjectivity or experience of oneself from a position of selfness, as I argued in Chapter 4. This also demonstrates both how robust the sense of self is and how fundamental or primitive self-consciousness is to the way we experience the world.

By introducing a minimal synchronic self, Strawson has by-passed one of the main problems with establishing the existence of the self – diachronicity or continuity over time. Accordingly, selves are real; they exist as a concrete entity, but they have only fleeting existence as a singular momentary self. This means that Strawson does not have to establish determinate identity conditions for his selves over time. Each one is unique and differentiated from the others by its temporal and spatial location. Each subjective experience of selfness will have its own unique neurological signature. In addition, Strawson claims that these selves have strong unity, 'strong-activity unity' such that they can rightly be judged objects. He concludes his case by stating that,

'...if there are any true unities in nature other than the unity of the universe, if there is an irreducible plurality of true unities in nature, then there are no better candidates for being such unities than thin-subject SESMET selves.'<sup>49</sup>

Strawson provides a convincing case for treating his SESMET-selves as objects. They appear to satisfy the most stringent conditions of objecthood. However, I am not convinced that they can count as selves on their own and in isolation from other parts of the cognitive system. There are also a few inconsistencies in his approach which I discuss below.

## **6.8 Strawson's selves: points of disagreement**

Strawson claims that consciousness is the hallmark of selfness. No consciousness, no self. One has to be aware to be aware of oneself. At the same time, if there is no awareness/consciousness/experience then there is no self. Many other major philosophers hold a similar view.<sup>50</sup> However, consciousness is also considered to be gappy. Dennett, in particular, used the gappiness of consciousness to deny the existence of the Cartesian self. Strawson uses it to defend his minimal multiple selves\* or SESMETs. According to Strawson, life is a string of broken moments of

---

<sup>49</sup> Strawson, 2009, p. 423.

<sup>50</sup> I have discussed several of these in the course of this thesis but see Dennett, Gallagher, DaMasio, Shoemaker, Nozick.

experiencing oneself as a subject of experience, even when some of those moments overlap. He claims that actual, in-the-present consciousness, only lasts up to 3 seconds (although in *Selves* he is reluctant to quantify the lived moment). This means the self only lasts as long as that experiential moment, however long that moment is. It also means that the self is only its experience of itself. There is nothing more to the self than these subjective moments, no matter how many of them there are.

Someone like Parfit could argue that the minimal momentary self is not a self at all, it is just a mental event.<sup>51</sup> All there is is an activation pattern in the brain. One could claim that the activation pattern just represents a particular state of the (cognitive) system. There is nothing to indicate it is a self-moment. There is nothing to indicate the presence of a self. Strawson defends himself against this view. He argues that this is what a self is. What makes it a self as opposed to just the experience is the fact that it is subjectively experienced, someone/thing is conscious of having that experience. This means they can make 'I' statements. He re-iterates the object/property distinction to affirm that experiences and experiencers are inseparable. While I accept his point, the difference between himself and Parfit appears minimal. If we put aside the object-status of the experience, one could claim that all Strawson does is include subjectivity. While this helps 'own' the mental event in that it is some one's experience, it is only owned by that one momentary self who is inseparable from that experience by definition. There seems little difference between positing multiple momentary experiences or positing a series of single mental events. One could say that the only difference between the two accounts is that one claims the mental event or experience is a self and the other claims it is just an event.

According to Nozick, the capacity for self-reference is both a necessary and sufficient condition for something to be a self.<sup>52</sup> The reason for this, in Nozick's account, is that the self does not exist prior to that act of self-referral. It is the act of reflective self-reference that brings the self into existence, in the having of an 'I' thought. The self comes into existence with the 'I' thought. This appears to match Strawson's position, and Strawson frequently refers to the similarities between their positions throughout his book. However, Nozick deliberately constrains his account

---

<sup>51</sup> Parfit argues exactly this in *Reasons and Persons*, Clarendon Press, Oxford, 1984.

<sup>52</sup> R. Nozick, *Philosophical Explanations*, Clarendon Press, Oxford, 1981.



of the self to awareness of oneself as an I or to the use of 'I' because he does not believe in a self as an independently existing thing. The self is a capacity for 'I' thoughts, nothing more. As discussed in Chapter 2, Nozick discounts the self because it lacks determinate identity conditions. Strawson tries to get round this by positing a string of momentary selves. Velleman puts forward a similar model which has selves spread out over time. He draws parallels between Lewis' 4D view of time and its impact on the status of objects, and multiple time-slice selves.<sup>53</sup> He claims that our perception of the self as a single entity always fully present at any given moment does not represent reality because the self that is present at any one time is just a momentary time-slice self, a self in that moment.

As I stated in Chapter 2, the implications of the 4D view of time hold equally for any object. Whether the 4D or the 3D view is true or not should have little impact on the identity conditions of the self. Both views have problems when it comes to objects that undergo change. All things, including selves, will be multiple time-slice entities. Selves would have no more nor less continuity than any existent concrete object can lay claim to. In contrast, and unlike bodies and other such objects, Strawson's selves lack continuity because they are *physically* discontinuous with each other, even when those self-moments overlap. There appears to be nothing linking one self\* to another self\* that would make it continuous over time in the way an object spread out in time would be.

Strawson claims that the phenomenology does not support a diachronic self, hence selves are synchronic only. There are two points to make here. First I don't think we experience our selves as continually recreated moment by moment, merely linked by the continuity of our bodies or our memories. That is not the standard account of our phenomenology. In addition, I think we do have a temporal sense which results in a sense of continuing through time, of things persisting in time and our selves continuing in time. Of course, just experiencing something a certain way does not make that the case; just because I experience continuity doesn't mean I am

---

<sup>53</sup> Velleman, 'So it goes,' *The Amherst lectures in philosophy: Lecture 1*, 2006, <http://www.amherstlecture.org/> See Lynne Rudder-Baker for a critique of 4D views of time where she argues that mereology is inadequate as an account of ordinary objects. 'A metaphysics of ordinary things and why we need it,' *Philosophy*, Vol. 83, 2008.

continuous. However, and in response to Strawson, I can equally claim that just because I *don't* experience myself as continuous doesn't mean I am not.

Second, Strawson says that our folk belief in diachronic existence is inferential. As such, it does not constitute our sense of the self and so can be discounted. I disagree with this conclusion. If there is no direct experience of being temporally extended, which I don't think is the case, then the idea of the diachronic self is, presumably, inferred from the evidence around us of temporally existing objects which can be accurately tracked through time and space. When this is not possible, we can usually work out why. We have broad experience of a range of different objects, organic and inorganic, artefactual and abstract, on which we base our judgements about what exists and how they exist. If our claims are justified and continue to be reliable, they move into the realm of what counts as knowledge. Why should our inferences about our own diachronic nature not be considered just as reliable as those we make about other objects in our world? The only time we encounter a minimal 3 second self is in fiction or academic texts. Serial selves are generally considered to be either drug induced or the product of neurological disorders. The idea of a continual agentic self may very well be an inference to the best explanation. It is one that seems to fit with the kind of world we live in.

Brook, in his critique of Strawson, argues for unity of consciousness as well as singularity.<sup>54</sup> He claims that singularity and unity are not the same. Singularity here represents a single experience which implies a singular consciousness. However, if one could be conscious of more than one thing at a time (a debatable point but seemingly supported by Brook's breakdown of the phenomenon of reading a book) these experiences would be multiple but united in that they are had by the singular consciousness or self. In this instance, unity does not rule out singularity and is compatible with it. In fact, what unites the multiple experiences is that they are being had by the one person and those experiences are inter-related. One may not be simultaneously experiencing oneself sitting in a chair and scanning the text but one's multi-experiences come together to form a whole experience, of reading a book.

---

<sup>54</sup> Andrew Brook, 'Unified consciousness and the self,' S Gallagher and J Shear (eds), *Models of the self*, Imprint Academic, UK, 1999.

Brook puts it succinctly, ‘synchronic unity at the level of phenomenology requires synchronic singularity at the level of ontology.’<sup>55</sup>

Strawson says there is no diachronic singularity, therefore there is no diachronic self. Given singularity is an essential feature of being a self, this would mean there is no continuous self. According to Brook, synchronic singularity implies synchronic unity. While Brook agrees with Strawson that there is no diachronic singularity (persisting as the same self over time), he is not so sure this implies there is no diachronic unity. If it doesn’t, there may be diachronic unity and it may be enough to establish a self over time, if unity is also considered an essential feature of any self. Brook puts forward the suggestion that diachronic unity of some kind is essential for selfhood as it would be hard to imagine a viable self that really was incapable of even the weakest form of future thinking. So there needs to be some sense of personal continuity. However, Brook only claims that the phenomenology needs to be present, as in a sense of continuity, even if there is no real continuity. Using Strawson’s argument, a continuous sense of self would lead to there being a continuous self.

Another concern is Strawson’s emphasis on purely self-reflective experiences as this places a severe constraint on what counts as a subjective experience. There are countless experiences that we have as conscious human beings that we would want to claim involve our selves, but which do not always involve overt self-reflection or ‘I’ thoughts. Protracted actions where one is engrossed in the act itself may lack overt self-reflection but are *self*-absorbing and require continuity of intention and attention. Towards the end of his book, Strawson appears to separate his position from Nozick’s in this regard and states that ‘a self exists during any conscious episode, not just during a fully self-conscious experiential episode.’<sup>56</sup> However, this claim is not consistent with the majority of his text, where he seems to require experience to be of oneself as an experiencing subject. It does not fit his definition of the SESMET which is a strongly self-reflective model. Subjectivity is awareness of oneself *as a self* experiencing or being a subject of experience. So there is an inconsistency here.

---

<sup>55</sup> Brook, 1999, p. 45.

<sup>56</sup> Strawson, 2009, p. 351.

In a similar vein, Strawson uses the human capacity to be engrossed in an activity for a period of time as an illustration of *loss* of self. This, he claims, highlights the gappiness of consciousness and the need to really focus on one's phenomenal experience. It is not a stream but a pulse. This indicates that Strawson is focussing on a particular kind of experience, one in which one is aware of oneself as a perceiving subject caught in the act of perception. Again, this parallels Nozick's version. However, one is not non-conscious during prolonged periods of absorbed action, or lacking in experiences or perceptions, both of self and other objects. When one is driving in an automatic way or reading an engrossing story, there is still an awareness of things other than the act of driving or the act of reading. If, as Strawson claims he does, one allows experience per se, then self-moments become more frequent and more extended. Consciousness becomes a lot less gappy and self-moments become much more densely packed. This still doesn't in itself establish diachronic continuity of selves but it does open up the possibility of a much more continuous experience of subjectivity, where explicit self-reflection is just one kind of subjective experience. It makes it more likely that we experience that very real sense of continuity, rather than just infer it.

In contrast, Nozick's condition for selfhood is clearly limited to explicit 'I' thoughts and so would rule out a broader conception of experience. At the same time, using 'I' thoughts as a determinant of self awareness is problematic. It assumes that there is no sense of self prior to that use, which would rule out young children, the cognitively impaired and other primates from having selves.

Strawson clearly indicates that patients with dementia have subjective experiences and says they see things from a position of selfness, so he could not support this outcome. A focus on 'I' thoughts also leads to the problem of how we learn to use the indexical appropriately without leading to circularity.<sup>57</sup> Nozick's account of the self excludes features that Strawson would want to include.

---

<sup>57</sup> See C Rovane, 'The Epistemology of First-Person Reference', *The Journal of Philosophy*, 1987, Vol.84, No. 2, pp. 147-167 and J Bermudez, *The Paradox of Self-Consciousness*, MIT Press, Bradford, 1998 for a good account of the problem.

Wilkes raises a related concern.<sup>58</sup> She claims that there are things about herself that are non-experiential, not conscious, or are no longer conscious. She has a fairly broad canvas and seems to include all the things (tacit or implicit beliefs and desires) that may influence who I am or how I can be. We do talk about finding out about ourselves or of not understanding ourselves. So there are self-features that are not always conscious or readily available to awareness, or that once were conscious but are no longer in the forefront of our conscious mind. This is one of the reasons why the term ‘person’ or ‘personal identity’ is not always synonymous with the term ‘self.’ There is a sense in which we feel there is more to our self than we reveal and there is more to our self than we know. We also talk about our ‘real’ self or ‘true’ self with the idea that it may be different from the one we currently manifest, points I referred to in Chapter 3. As I discussed in that chapter, some of these items may be more closely tied to the issue of personal identity and, as does Strawson, I draw a distinction between self functions and personal identity. Having said that, the self as I would define it, is not limited to just those things in consciousness. Contrary to Strawson’s claim, I would argue that the self is distinct from the sense of self in that the self creates the sense of self and there may be non-conscious/no longer conscious influences at play. Strawson has, however, collapsed the self into the sense of self. There is nothing over and above the experiential moment. In this, he ignores the very mechanisms that make it a self-defining moment, which I address below.

## **6.9 The momentary/minimal self is not enough to be a self**

My main concern with Strawson’s account of the minimal self is its viability. Strawson has approached the problem of the self (or selves as he prefers) from a phenomenological perspective.<sup>59</sup> Using this method, he narrowed down the essential features of selfness to singular subjectivity. A self, he claimed, would still exist if there were just that moment of lived experience. That moment would be instantiated by a particular pattern of neuronal activation. That is what it is to experience oneself as a self, to experience an ‘I’ moment, regardless of any other past ‘I’ moments. This moment in time, the lived present or NOW is independent of all other such moments.

---

<sup>58</sup> Wilkes, ‘Know Thyself, 1999.

<sup>59</sup> Zahavi and Parnas argue that Strawson misuses the term ‘phenomenological’ and his is not a genuine phenomenological approach. D Zahavi and J Parnas, ‘Phenomenal consciousness and self-awareness: a phenomenological critique of representational theory,’ *Models of the Self*, editors S Gallagher and J Shear, Imprint Academic, U.K., 1999.

The self that is its subject has only come into existence at this moment. It is unique and it is brand new. It did not exist before and will not exist again. It is self-contained and object-like, singular and unified, a distinct identifiable item with a specific well-defined physical instantiation.

This self does not seem viable *as a self*, given what we know of the brain, perception and current neuro-physiology. Experiencing oneself as oneself while in the act of perceiving or experiencing something other than oneself involves a complexity of processes that are still not fully understood. First, and in line with Strawson's model, in order for this experience to count as a self-experience, and hence a self, it has to have certain properties. It has to be experienced in a very specific way. It has to be experienced as singular, it has to be experienced as mental, it has to be experienced as subjective. For it to be subjective or experienced from the subject's perspective, it has to be a conscious experience. For it to be about a subject, there must be some awareness (possibly tacit or even innate knowledge?) of what it is to be a self. There has to be contained in the experience, an awareness of one's own singularity and distinctness from everything else that surrounds one. There has to be a sense of 'this is me, whatever me is.' In effect, there has to be self-consciousness. This experience also has to preclude a purely bodily experience because a Strawsonian self is purely mental. This is experience of oneself as a mental something, an inner subjectivity. That is a lot of differentiation or cognition going on about a range of very different things.

In addition to that, there is the content of the experience itself. Here I am not making a conceptual distinction between the experience, the experiencer and the content of experience mentioned by Strawson. My point is merely the obvious one that the experience, first and foremost, is of something (or about something). So the SESMET or self-experience has to represent whatever the content of that experience is. There will be the perception of that experience (or the mental representation of that experience) and possibly the sensation of that experience (the quale or what it is like). In addition, there has to be not only cognition of that content (awareness of what it is) but also cognition of its relationship to you, the perceiver. The experience will be from a clearly defined perspective or point of view that provides information of one's location in space and time, one's current state and one's response to the

content of the experience. The experience will have meaning or significance to you, the observer or experiencer, that provides you affordances. These things are minimal to there being an experience of anything from a position of a subject, in order for it to count as a moment of selfness (i.e. to use Strawson's distinction - me\* experiencing this right here and now as a conscious occurrence of me\*).

It is difficult to imagine how the singular neuronal response, identified by Strawson as a strong unity such that it counts as a singular object, could present all the information just mentioned such that it would register as a self-moment, a moment of lived experience *in isolation* from other neuronal inputs both past and present.<sup>60</sup> Neurophysiology is only just starting to come to grips with what it is to be conscious and the underlying neural processes of consciousness are still under contention. It is not known why some things register in consciousness and others don't.<sup>61</sup> Some hold that there is a threshold, above which phenomena will register as conscious; others that there is a global workspace through which mental items need to pass to become conscious.<sup>62</sup> More recently, there appears to be evidence that consciousness is distributed such that it is difficult to pinpoint the mechanisms responsible for what is or is not registered as conscious. There are also multiple perceptual processes involved in any experience which make it difficult to see how it could be narrowed to one particular neuronal cluster. To make such an experience an 'I' moment requires the body/brain to have an already robust sense of itself as an individual thing. While this becomes more refined and developed over time, research indicates that this process is already occurring in very young infants.<sup>63</sup> Even at this level, the patterns of activation needed to signify to the infant that 'this is happening to this body' are very complex. This would put in doubt Strawson's claim that an experience of oneself as experiencing something can be explained by recourse to just one set of neurons at one particular time.

---

<sup>60</sup> J O'Brien and J Opie put forward a connectionist model of individuated conscious thoughts which are characterised by 'stable (versus unstable) patterns of activation in neural networks' that could theoretically represent that thought even if isolated in a glass of water. This view is susceptible to similar criticisms to those above. See 'Multiplicity of consciousness and the emergence of self' in *The Self in Neuroscience and Psychiatry*, eds. Tilo Kirscher and Anthony David, CUP, 2003.

<sup>61</sup> See O Flanagan, *The Science of the Mind* (2<sup>nd</sup> edition), MIT Press, Cambridge, Mass., 1991.

<sup>62</sup> Dennett critiqued this idea in *Consciousness explained* but it is the cornerstone of Metzinger's account in *Being no-one*. See Chapter 5, 2003.

<sup>63</sup> M Legerstee, 'Mental and bodily awareness in infancy: consciousness of self-existence, *Models of the Self*, editors S Gallagher and J Shear, Imprint Academic, U.K., 1999.

It is more than likely that the one singular unified ‘self’ moment is reliant on a web of other neuronal and bodily processes happening simultaneously for it to be experienced as a ‘self’ moment. While any given ‘self moment’ may be definable, even reducible, to a single neuronal cluster, it is unlikely that that cluster will register as such without being embedded in a much larger network. In fact, it is highly likely that there will need to be past experiences and genuine continuity in order for that experience to be both an experience of something particular and one owned and experienced by this particular identifiable consciousness - me. It is also highly unlikely that a single neuronal cluster will produce the kind of robust phenomenal experience that we are familiar with when we have such a moment of self-recognition in isolation. Even a minimal self, lacking knowledge of its identity, with no sense of place and time or other emotional connections, can experience intense delight or suffering that would still require a rich array of perceptual apparatus and other neuronal connections for it to be experienced as such.

Strawson’s minimal self should really be called a momentary self. A minimal self is still a viable self, despite the removal of some of its characteristics. However, continuity of some more durable kind would seem to be essential to turn a moment of subjectivity into a minimal self experience. In my final chapter, I argue that a system of specific self-mechanisms is needed to create the phenomenology requisite for a minimal self-experience. These mechanisms are unified enough by the principles around which they are organised, to count as an object, what one could call a self. I also suggest that consciousness, as a primitive feature, can provide a phenomenal sense of continuity that is essential for a self to be more than moments of subjectivity.

## **6.10 Phenomenology and the experience of the body**

Strawson underplays the role of the body and what he calls the EEE perspective (environmentally embedded, embodied, ecological, enactive). The self-experience, he claims, is not identical to the body. While Strawson doesn’t deny bodily experience, agreeing that interoception and somatosensory awareness are the foundation of self-experience, he claims that we experience the self as a ‘mental’ thing predominantly. There is, he claims, no competition. Not all experience is sensory and we are far more aware of our (often non-sensory) thinking than we are



of anything else. He says the distinction is between ‘focussed, express, attentive, thetic awareness, on the one hand, and unfocussed, more or less dim, non-attentive, non-thetic awareness on the other.’<sup>64</sup> Humans have a lived awareness of the privacy and secrecy of their inner, mental world. This defines their self-experience and what separates it from their experience of themselves as an embodied human being. Thus, for Strawson, embodiment is not the hallmark of self-experience. It is the mental that defines self-experience and which leads to the sense of a self distinct from one’s whole person. In taking up this position, Strawson appears to reduce the self to the ‘thinker’ of the Cartesian model or to the ‘I’ or knower of the Jamesian model.<sup>65</sup>

This limitation, as I have already discussed, is problematic. Even though Strawson incorporates other bodily sensations as part of the self-experience, he ranks them as minimal. They don’t impact on the self-experience or the sense of self. In his section on the relationship between self-experience and self-consciousness, Strawson reiterates the mental nature of self-experience. True self-experience is not just experience of oneself as a self but ‘experience of oneself *expressly grasped as oneself* (his italics).’<sup>66</sup> This is a strong claim and seems to restrict self experience to the problematic self-reflective element mentioned above. On the other hand, Strawson does not want to rule out young children from having self-experiences and in a later chapter claims that one can be a self as long as one is a *subject* of experience. Despite this qualification, there does seem to be some equivocation between these points. He wants to draw a line between just having any experience and experiencing self, because snails and fleas as well as dogs and cats could then be a self, if they can be said to have experiences of some kind.

It may be the case that the (conscious) experiences of a wide range of non-human creatures do not include a sense of the mental or a sense of the singleness of one’s self, although this can’t be ruled out completely. Yet by making reflectiveness the defining feature of selfhood, he is in danger of ruling out selfness in some humans, especially young children and humans with cognitive impairment or neurological illnesses. One could suggest that the experience of some kind of subjectivity is

---

<sup>64</sup> Strawson, 2009, p.30.

<sup>65</sup> James divides the self into two primary categories, the self-as-object (the ‘me’ self) and the self-as-knower (the ‘I’ self). W. James, Chapter 10, 1890, p.239.

<sup>66</sup> Strawson, 2009, p.107.

possible in some animals and immature or damaged humans but that selfness requires reflection. However, Strawson has already collapsed being a subject of experience into being a self. In addition, making reflectiveness the defining feature ignores the fact that many, if not all, of our non-reflective experiences and perceptions are, nevertheless, self-defining and self-informing, a case made strongly by Gibson and Bermudez.<sup>67</sup>

Strawson could argue that sensation-type experiences are about the whole human being and serve to identify and individuate one's body from the world but not one's self. They do not, in themselves, count as a self-experience. So if one is experiencing something and is just aware of undergoing that experience in a non-reflective way, then this does not appear to count as 1) self consciousness or 2) subjectivity. Strawson claims the self in these instances is invisible. Thus, it is not mentally present and not part of the self-experience. However, as mentioned above, one could argue that all conscious perception is self-defining and self-individuating, even if the (bodily) sensations are unnoticed. There is research evidence to show that the loss of these seemingly invisible sensations does impact on one's sense of self, indicating that they play a bigger role than is apparent when we self-reflect.<sup>68</sup> Humans have particularly refined mechanisms for differentiating self from other, including their internal states. If these experiences are self-defining and conscious, then they are also subjective. They are about 'me.' There seems no good reason to claim that they should be excluded from contributing to the sense of self. Acknowledgement of their role would give a more robust sense of self that could incorporate stronger continuity conditions and contribute to a phenomenological sense of existing over time, at least for longer than 3 seconds.

Again, this highlights the difficulties inherent in using an analysis of one's own self-phenomena to drive one's model. For example, I would argue that our self-phenomenology includes a sense of our self as being 'inside' our bodies, even if at times we are at one with that body making it invisible and at other times completely

---

<sup>67</sup> I have discussed Bermudez in earlier chapters, see *The Paradox of Self-Consciousness*. I discuss Gibson in more detail in the next chapter. Edward Reed, and Rebecca Jones, (eds), *Reasons for Realism: selected essays of J.J.Gibson*, Lawrence Erlbaum Assoc., Hillsdale, N.J, 1982.

<sup>68</sup> See O Sacks, *The Man Who Mistook His Wife for a Hat*, Summit Books, USA, 1985 and V Ramachandran and Sandra Blakeslee, *Phantoms in the Brain : Probing the Mysteries of the Human Mind*, 1998.

alienated from it. This condition does not lead one to assume that one's self is identical to that body, precisely because the experience is of an *inner* sense. It contributes to the mental view of the self because it incorporates a strong sense of being on the inside looking out. As with the Cartesian 'inner homunculus,' it is easy to objectify the body for this reason. Thus, Strawson may well be wrong when he says that we don't directly experience our self\* as embodied. In fact, having that sense may lead to an *experience* of continuity, that one's self does not cease in between moments of overt self-reflection.

Zahavi and Parnas take issue with Strawson's claim that he is starting from the phenomenology of self.<sup>69</sup> According to Zahavi and Parnas, his is not a true phenomenological position because Strawson is working within a tight physicalist analytic framework. They claim Strawson tries to explain the phenomena of the sense of self from an objective scientific position. Zahavi and Parnas say this is the wrong way round. The phenomenological approach is to lead you from the position of selfness out into the world to explain how you get objectivity and intersubjectivity through that process. If Strawson were to use a true phenomenological approach then he could not ignore the property of ipseity, the fundamental grounding of self-awareness. According to the Continental tradition, while the self is given in phenomenal experience, it is non-reflective and non-inferential. This self-giveness is the invariant feature that provides continuity to all consequent experience. Memories, for example, are aspects of this continuity because they are woven by you into a fabric, not just remembered by you as objects. I agree with Zahavi and Parnas that Strawson does not adopt a true phenomenological position. In his defence, this is because he is not using it as an 'epoche,' a methodology for exploring self-world relations.<sup>70</sup> He accepts the phenomena and, in true analytic style, tries to uncover what is behind or underlying the phenomena that may cause it to occur. Where he differs from other physicalist accounts like Dennett's and Metzinger's, is that he assumes the phenomena is real and an accurate portrayal of what it is to be a self. He accepts the phenomenology and uses it as the basis of his model. This is an advantage. However, his minimal self ends up being not much of a self at all.

---

<sup>69</sup> Zahavi and Parnas, 1999.

<sup>70</sup> Husserl uses the phenomenological approach as the framework for discovering how the world is knowable. See i.e. Poul Lübcke, 'A Semantic Interpretation of Husserl's Epoché,' *Synthese*, 1999, Volume 118, Issue 1, pp 1-12.

## 6.11 Discontinuous or continuous selves?

Strawson's account of what a self is raises several related issues. The first is whether or not self-experience, as described by Strawson, is as momentary and disconnected as he argues it is or experiences it to be. I would argue that it isn't. Because he collapses his sense of self into the self, he can only examine what manifests in his own phenomenology. That becomes the totality of what a self is. I would argue that the self is that (system) that enables the experience of a rose, for example, to simultaneously be an experience of self. It is whatever creates that phenomenology. At the same time, the same system is responsible for much more than this minimalist sense of self.

A second related point, dealt with briefly above, is whether the self is experienced as purely mental rather than embodied to some extent. This point impacts on Strawson's claim that continuity is not experiential and selves are only momentary. It is my contention that self-experience incorporates an awareness of being embodied, precisely because we know immediately when it is weakened or different from the normal experience. We know when things are wrong, even slightly. In addition, the sensory modalities of the body, along with its internal movements and rhythms, create a background 'hum' that is always at the periphery of one's consciousness. This bodily hum could act as a continuous context for those explicitly self-reflective moments.<sup>71</sup> The third point is whether or not a momentary self is long enough to count as a self. How much diachronic continuity do we need for self-moments to be considered moments-of-the-same-self? I argue that the Strawsonian self does not adequately address the robustness of our self experience or our self concept and thereby, does not reflect our lived reality. I also contend that a sense of ourselves as players in the world, as beings who can make and enact plans, relies on a core belief in our own *self* continuity, not just that of our bodies and brains. Just because we have difficulty accommodating the notion of change into our ideas of identity and sameness, does not mean that the same thing cannot change and yet remain itself.

Strawson says that those who advocate for a (real) self assume it has to last a lifetime. It has to be the same immutable self from birth to death. There is, however, a

---

<sup>71</sup> I deal with this in more depth in the next chapter. See Heller-Roazen's *The inner sense* and also Damasio, *Descartes' Error*, Avon, New York, 1994 for an explication of this sense.

huge divide between a life-long same self and one that lasts less than half a second. If I want to demonstrate that there is such a thing as a self, I don't need that self to last a life-time. I don't even need it to last half a life-time. I would, however, need it to have the *potential* to last as long as the body in which it inheres. But I could even accommodate the existence of serial selves as in cases of DID, if each self had enough continuity to count as a distinct identifiable and re-identifiable self. To draw a parallel, we do not demand that all our every day objects last 'a life-time,' whatever that may be, in order to consider them real, concrete objects.<sup>72</sup> Some things have short life-spans compared to other things. It is inevitable that some things will come into existence only to be destroyed moments (even seconds) later. So if selves exist, and this is what Strawson is arguing for, there will be a *potential* life-span for that self, barring the kinds of things that can intercede and shorten that life-span. The difference between my account and Strawson's is that my selves could, potentially, last a lifetime however long that may be. Strawson's selves could not.

Strawson argues against a system view of the self because it fails the strong-unity object test. However, that unity test is only one particular view of what constitutes a singular object. As he says, humans, horses and a whole range of other objects fail that test as well. It is likely that the system view could satisfy the object test if the requirements were a little less stringent, as outlined at the end of Chapter 2. More importantly, Strawson's selves are identical to the experience of the self-moment or *syssele*. There is nothing outside that moment. This means selves can't exist past that moment as the same self. The only way a self could exist for longer and be (comparatively) continuous would be if experience were dynamic, rather than discrete. Even then, as with Locke's account, it could not survive the gaps, the periods of non-reflective consciousness. With a system view, the self is not identical to the self-phenomena. The system-self creates the phenomena. This enables the self to 1) continue if the mechanisms are working and 2) contain more than the content of current experiences. This could address Wilkes' criticism that her 'self' contains more than just her current experience.<sup>73</sup>

---

<sup>72</sup> The issue with diachronic selves is which bit of them remains the same such that we can say it is the same self over time and change. My point is that this is a problem for any developing life form and is even a problem for objects like Rudder-Baker's bronze statue or the Ship of Theseus.

<sup>73</sup> Wilkes, 'Know Thyself,' 1999, pp. 25-38.

Strawson argues for a minimal self because it is the only kind of self that he thinks is defensible and that accords with the phenomenology. I am not convinced it does accord with the phenomenology. I also think that we should not put the self outside of our normal epistemological framework. Our experience of all objects in the world is of discontinuous instances. As Strawson states when arguing that consciousness is gappy, our eyes saccade, there are blindspots and we fill in details not present to perception. None of this makes us form the conclusion that objects flit in and out of existence or that they are not continuous with themselves over time. We infer that they have diachronic continuity. We could, legitimately, make the same claims about the self. However, this would require that the self not be reduced to just those properties that Strawson has identified as essential. The self would have to be more than a sysele or a SESMET.

In the next chapter, I address the phenomenology of the self in more detail to establish the viability of a continuous self. In particular, I look at Damasio's model of the self as a viable alternative to both the narrative self model and Strawson's minimal self. Damasio, like Strawson, claims that the self exists. I also discuss the plausibility of a system-self, based on current neurological findings that seem to identify mechanisms that play critical roles in forming our sense of self and which seem responsible for our self-phenomenology. I claim that such a system could be considered object-like and concrete. It could also count as an alternative to the Cartesian self.

## CHAPTER 7

# DAMASIO AND THE SELF AS A SYSTEM

My self-being, my consciousness and feeling of myself, that taste of myself, of I and me above and in all things, which is more distinctive than the taste of ale or alum, more distinctive than the smell of walnut leaf or camphor, and is incommunicable by any means to another man...Nothing in nature comes near this unspeakable stress of pitch, distinctiveness, and selving, this self-being of my own.<sup>1</sup>

### 7.1 The importance of the phenomenology of the self

In the preceding chapters of this thesis, I have tried to do several things. In Chapters 1 and 2, I argued that many of the reasons given to demonstrate that the self is not an independent entity or object apply equally to many everyday concrete objects that we ordinarily assume exist, such as cars, teaspoons or even apples. Many of these everyday objects have difficulty satisfying criteria for ontological existence or for determinate identity. In certain circumstances, it will be indeterminate whether or not they can be considered the same object over time and change, as demonstrated by the Ship of Theseus example. In these respects, the self is similar to all seemingly existent things. If one claims that the self exists but that it is not real or it is an abstract entity, one has to do more than just show that it lacks determinate or diachronic identity. In Chapter 4, I extended this argument to demonstrate that the pathologies of the self only demonstrate that the self can be damaged or destroyed like other existent objects. In itself, this is not an indication that there never was a self to start with or that the self there was not real. If nothing else, the debilitating effects of self disorders indicate how important the self is to well-being. The fact that the self can be damaged, or even destroyed, would indicate that something had changed, that what was in existence before is now no longer. As we know, all objects in our world can be damaged or destroyed. These are not sufficient grounds to discount their existence. The same is true for the self. We should not demand invulnerability from the self, given we do not demand invulnerability from our turkeys or teaspoons in order to postulate their existence.

---

<sup>1</sup> Gerard Manley Hopkins in 'Notes,' *The Poems of Gerard Manley Hopkins*. Ed. Robert Bridges. London: OUP, 1930, p.123.

In a similar vein, and as discussed in Chapter 5, there is a tension between accepting the claims we make about the world we live in and denying the reliability of the claims we make about ourselves. If our self-phenomenology is so mistaken and unreliable that we cannot use it to support the claims we make about what it is like to be a conscious being, then we would be silly to rely on the same processes to support the claims we make about other objects we perceive. While it is true that the claims we make about external objects can be verified by others, we all share the same perceptual apparatus. As we know from experience, we make similar mistakes and suffer the same illusions.<sup>2</sup> Nevertheless, we still come to reliable conclusions about the world and share a similar perspective on how things appear to us. My point with all this is to indicate that there needs to be some ontological and epistemological consistency. When it comes to the existence or not of the self, the ‘evidence’ bar appears set at a different level. The self-as-object has to satisfy a much more stringent set of criteria than we demand of other objects. While the identity criteria for selves may well need to be different from those of other objects, as Locke and Lowe postulate, it is not clear that they should be so far removed from those of ordinary objects that they are impossible to satisfy.<sup>3</sup> It is highly unlikely that all our self-experiences are fallacious, illusory and based on nothing. Our predominantly shared phenomenology would seem to indicate that our reports on our self-experience pick out real features of that experience.

One important feature of the self that seems to set it apart from other familiar objects is the subject/object distinction. The self seems to lack any object-like physical characteristics, such as weight, mass and location. At the same time, it has robust subject-like characteristics, as presented throughout this thesis and poetically captured by Manley-Hopkin’s quotation. Because we experience ourselves subjectively, from the inside (often referred to as a privileged position), it is claimed that, as an acting self, we don’t really experience ourselves as *a self*. We don’t experience or observe our *self* objectively, as an object. It appears to be the nature of the self that it is only experienced as a subject of experience. This causes some doubt about the existence of the self as an

---

<sup>2</sup> See Alan Chalmers, *What is this thing called science?* 2006.

<sup>3</sup> E. J. Lowe, *The Possibility of Metaphysics: Substance, Identity and Time*, Clarendon Press, Oxford, 1998.



object in the world and appears to be the basis of Hume's scepticism. As mentioned in earlier chapters, Cassam calls this view the 'elusiveness thesis.'<sup>4</sup> He claims that any materialist conception of the self entails awareness of self as object, not just as subject. Viewing the body as an object is not sufficient, unless the body is considered identical with self, a view I have not been supporting.

It has been part of my intention in this thesis to argue that the seeming *lack* of objective features of the self has not been unequivocally established, despite claims to the contrary. If we take the anti-self arguments as presented by Parfit, Dennett and Metzinger, they rest on the absence of either an immaterial self or of a singular neuron or other non-reducible entity existing in the brain or elsewhere that takes the role of the Cartesian self. As I argued in Chapter 1, a rejection of a Cartesian self does not mean there can be no other type of self; neither should positing the existence of a self depend on finding an irreducible singularity. Some other *complex physically-instantiated self* could count. It would not be unreasonable to suppose that the self-as-object is inferred in much the same way we infer the existence of other objects, based on their perceived characteristics. Our sense of self is representative of a widely shared phenomenology. It is that phenomenology that leads to the belief in a self. An analogous situation would be the positing of atoms based on the observations of their movement through a cloud chamber.<sup>5</sup> We experience the effects of the self or observe its actions and infer there is something that causes both.

The issue as noted is that it is not obvious that it is one thing, a self that is responsible for that phenomenology and those actions. Many of the authors referred to in this thesis argue that what we take as a singular thing, the self, is actually underpinned by multiple processes. As Hardcastle acknowledged in her review,

---

<sup>4</sup> See Quassim Cassam, 'Introspection and bodily self ascription', *The Body and the Self*, ed Bermudez, Marcel and Eilan, MIT Press, 1995.

<sup>5</sup> J.J. Smart argued that atoms are not instrumentalist fictions but real theoretical entities that are posited as existing. See *Philosophy and Scientific Realism*, Routledge and Kegan Paul, London, 1963.

Two points stand out ... (1) little has changed theoretically in the intervening decade, and (2) the contributors all agree that there is no such thing as a unified self. ...<sup>6</sup>

This latter point she claims is interesting theoretically, given the contributors ‘come from radically different backgrounds with radically different points of view.’ I have been attempting to demonstrate that this is not the case, that the self could be considered unified, even singular if there existed a structure located in the brain’s architecture with enough unity and internal cohesion to count as a singular system whose functions were self-maintenance, self-differentiation and self awareness. Such a structure could be considered a concrete object even if comprised of parts. It could be identified as a self if it could be shown that that structure underpinned the production of our basic self-phenomenology. It may be that such a self will not be experienced *as an object*, and its existence as an object may only be inferred from the evidence. However, I think such evidence would be convincing.

In Chapters 2, 3 and 5, I addressed the most common conceptions of the self, each of which place the self either as non-existent or, if existent, then abstract or illusory. As I claimed, the reasons appear to be ontological and reflective of the Cartesian constraints – such as the need for singularity or immutability (like the Dennettian ‘soul-pearl’). As a consequence, two of the most widespread and popular conceptions of the self claim either that the self is the same as the person or personal identity, or claim that the self is a narrative or self-story. I have argued against both these conceptions, claiming that the self is more than just a term of reference and more than a narrative, given neither position is capable of adequately explaining or capturing our self-phenomenology. Many of the self’s defining features, such as subjectivity, self-reflection and sense of agency, are not explained by focussing on personal or narrative identity alone.

In fact, I have tried to demonstrate that the self can be separated from its personal identity, by focussing on those facets of the self that remain largely invariant, such as its

---

<sup>6</sup> The contributors were Dennett, Neisser, Gazzaniga, Lewis and Keen. Taken from the introduction by Valerie Gray Hardcastle, ‘A Self Divided – A Review of Self and Consciousness: Multiple Perspectives’ Frank S. Kessel, Pamela M. Cole, and Dale L. Johnson (Eds.) *PSYCHE*, 2(1), April 1995 <http://psyche.cs.monash.edu.au/v2/psyche-2-01-hardcastle.html> .

individuating capacity, its continuity and numerical identity and, most importantly, its ipseity. I argued that these invariant features of selfhood become apparent in pathological cases where personal and narrative identity are lost but some form of selfhood remains. At the same time, the aetiology of many of the self-pathologies place the self and its phenomenology as emerging from specific neurological processes in the brain, rather than from the adoption of a social norm or the construction of a self-narrative. This evidence tends to ground the self in human biology and neurology and not purely as emerging from social interactions.

Interestingly, the neurological data has been variously interpreted and is often used to confirm the *non-existence* of the self (i.e. given visual perception is a distributed process). I have argued, most notably in Chapter 5, that the problem with these types of positions is that they assume that a viable self means a Cartesian ‘soul-pearl.’ If there is no singular self then the phenomenology is false. Most alternative accounts of the self work by dismissing the self-phenomenology and, hence, the reliability of self knowledge as evidence of a self. The self and its phenomenology become illusions. This leads, inevitably, to claims that there is no such thing as phenomenal consciousness. I have argued that one can’t just dismiss the phenomenology as illusory, given how universal such experiences are and how uniform their supervenience on specific functional neuronal clusters in the brain is. Strawson is a notable exception in that he uses the phenomenology as the basis to posit a self that is both real and concrete. However, as I set out in Chapter 6, the self he posits is minimal. It is purely synchronous and exists as a momentary period of consciousness of less than 0.3 seconds duration. This appears too minimal and momentary to account for the robustness of our phenomenal experiences. Against his view, I argued that even a momentary self would require a substantial network of processes to make it a *self-conscious* moment. A single neuronal spike of selfness is unlikely to account for the robust phenomenological experience of ‘being me at this moment.’ Hopkins’ robust self-phenomenon needs to be explained. None of the accounts I have addressed so far come close to explaining this sense of self being. At the same time, there is an onus on anyone establishing the existence of an object to demonstrate that it has enough coherence, individuation and unique identity to qualify as a single thing. As I set out at the end of Chapter 2, the self does need to be something

that can be re-identified as that same self through time, if not change. Experiencing continuity or needing continuity is not sufficient to show there is continuity.<sup>7</sup>

In this chapter, I want to address a theory of the self which I feel comes closer than any others in dealing with the object/subject dichotomy mentioned above. In *Self comes to mind*, Damasio puts forward his account of how conscious selves come into being<sup>8</sup>. Like others, his account is non-Cartesian; his interpretation of the way the neuronal structures work to produce our phenomenology is deeply rooted in neuroscience. Despite this, he claims unequivocally that selves exist. His description of the development of selves within the human/animal cognitive system proffers an account of an entity-like structure that underpins and accounts for the production of our self-phenomenology. I claim that such a closely bound network of self-systems could qualify as a singular but complex object. Contra Dennett, this would have mass and location, even if it were physically distributed.<sup>9</sup>

In the following sections, I give a detailed explication of Damasio's position. In the process, I hope to demonstrate the strengths of his model and why it seems closer to the way we are in the world than other current theories. I will also draw on the findings of other neurologists that appear to support his claims. I will conclude that the self-defining/self-referring structures in the brain work together as a self-system that 1) produces our sense of selfness and self-awareness, 2) produces our capacity to think 'I' thoughts and 3) enables self-generated action. Such a system is a candidate for classification as a complex object in its own right, much as the visual system is. It is as real as a teaspoon and is a candidate for a self.

## 7.2 Damasio and the beginning of the self

I have referred to Damasio's model of the self in earlier chapters of this thesis for two reasons. First, his theory of what the self is emerges from current neuroscience. It is

---

<sup>7</sup> This is a point raised by my supervisor, Ian Ravenscroft, in discussions.

<sup>8</sup> A.R. Damasio, *Self Comes to Mind*, Pantheon, New York, 2010.

<sup>9</sup> In personal correspondence, Dennett argued that the self lacked substance. He said it would be wrong to ask, 'How much does it weigh?' However, if it were a set of neurons or an activation pattern, it may still be problematic to ask this same question.

grounded in what we currently know about certain parts of the brain, what those parts seem responsible for, and what the effects on our phenomenology are should they be damaged or compromised in some way. His claims are not only supported by this scientific evidence but emerge from it. Second, he does not ignore the sense of self but uses the phenomenology as grounds to support his claims. In fact, he explains how the phenomenology is directly linked to what is occurring in the body. As a consequence, his explication of the self is neither Cartesian nor epiphenomenal.<sup>10</sup> Yet, according to Damasio, selves definitely exist.<sup>11</sup> This makes his approach novel.

It is interesting that Damasio comes to quite radically different conclusions to both Dennett and Metzinger, even though in *Self comes to mind*, he is exploring a similar link between consciousness, the mind and the self and, like them, assumes that each of these three items will be explicable by recourse to physicalist descriptions, most notably from what we know of the brain and biological systems through the findings of neuroscience, biology and psychology. Damasio, like Dennett and Metzinger, justifies his claims and conclusions by adverting to current findings in neuroscience and evidence drawn from self-pathologies. However, he interprets those findings in a radically different way. Rather than proof that selves are illusions, he argues that the evidence demonstrates that selves exist. I think this difference can be explained by where Damasio places the self and consciousness in the cognitive structure. He does not believe that consciousness or selfness are late biological acquisitions or are higher order properties of the brain. For Damasio, having a sense of self(ness) is fundamental to perception and it begins with the first perceptual discrimination by what Damasio labels a ‘proto-self.’<sup>12</sup> This places selves as fundamental features of conscious existence. Perception is self-constituting<sup>13</sup>

---

<sup>10</sup> His non-core self is mildly narrative because it is autobiographical. However, in the literature autobiographies are treated as less fictive than narratives and second, its emergence rests on the core self without which there could be no subjectivity and no autobiographical self.

<sup>11</sup> Damasio, 2010, p.8.

<sup>12</sup> At this level, the sense of self is more a sense of *selfness* as it is a primitive sense of just being aware of being. Damasio argues that this is likely to be shared by other sentient creatures.

<sup>13</sup> In this respect, Damasio’s model of perception resembles Gibson’s direct realist approach. I have referred to Gibson’s account of perception in earlier chapters. He claims much of the discrimination of objects is pre-cognitive and driven by real features of the perceiving object and subject, such as

Interestingly, Damasio reverses the customary explanatory order, arguing that the only reason we are conscious is because we have selves. Without a self, or more precisely a self process, there would be no consciousness. He claims that,

Conscious minds arise when a self process is added onto a basic mind process. When selves do not occur within minds, those minds are not conscious.<sup>14</sup>

Damasio does not classify mind processes or self processes as purely human attributes. Minds can exist, albeit in more primitive form, in all sorts of animals. So can primitive selves. In Damasio's account, the advent of (primitive) consciousness is the awareness of being, and this is the awareness of being an entity - a single, separate entity. This is the beginnings of consciousness of 'self.' It is also the beginnings of subjectivity. This means that a broad range of animals would be conscious or even self-conscious. This view is shared by other neurologists. Feinberg and Keegan, referred to in Chapter 5, place the self in specific areas of the brain, associated with identifiable properties.<sup>15</sup> Panksepp and Northoff argue that there is a lawlike core SELF common to mammals that, while primitive or primary, is the foundation or seat of the human self. This SELF is a physical brain process rather than a conceptual entity.<sup>16</sup> Like Damasio, they argue that there is a proto-SELF,

a poorly understood genetically provided complex network infrastructure, similar across all mammalian species, [that]is a necessary substrate upon which higher idiographic selves are epigenetically constructed.<sup>17</sup>

---

affordances and perspective. This is the mutually constitutive process that Merleau-Ponty describes in *Phenomenology of perception*.

<sup>14</sup> Damasio, 2010, p.8. There are several ontological implications in this statement which have a bearing on my thesis. One is that he views the mind as a process as well as the self. Hence, both will have the same ontological status. If the mind is considered an entity then so should the self. If the mind can be considered an object, then so should the self. Likewise, if not, this goes for both. At the same time, the *reality* of the mind or the mind-process will analogically transfer to the self or the self-process. Thus, if minds are considered to exist, so should selves.

<sup>15</sup> Feinberg and Keegan, 'Where in the brain is the self?' *Consciousness and Cognition*, Vol. 14, 2005, pp. 661–678.

<sup>16</sup> Jaak Panksepp and Georg Northoff. 'The trans-species core SELF: The emergence of active cultural and neuro-ecological agents through self-related processing within subcortical-cortical midline networks,' *Consciousness and Cognition*, 2009, Vol. 18, pp. 193–215. The authors explain the use of SELF as follows: 'The use of capitalization has many functions, but the primary one is to highlight that we are specifying an actual brain process as opposed to a conceptual entity' (p.194).

<sup>17</sup> Panksepp and Northoff, 2009, p. 194.

They posit a ‘core set of interrelated structures involved in creating the basic sense of SELF.’ Damasio claims that a brain can only be conscious if there is subjectivity. Subjectivity requires the brain/mind process to be aware of itself as an entity that is experiencing something, triggered by engagement with an object. The first object is its own body and bodily sensations. If it is aware of experiencing something, then it is aware of itself through a kind of reflexivity. One entails the other. Thus self and subjectivity are the same thing. Again, this view is reflected in Panksepp and Northoff’s paper on the core SELF. While they do not say that self precedes consciousness, they acknowledge the inseparable link between the two. They also place phenomenal consciousness and selfness at a foundational level, rather than as a higher-order property. This self-entity also emerges long before we get persons or personal identity.

At this pre-reflective, pre-cognitive and pre-linguistic level, phenomenal consciousness and pre-reflective self-instantiation probably go hand in hand, being inseparably linked to each other. This implies that the occurrence of some kind of basic SELF in the sense described here is linked to the occurrence of foundational phenomenal consciousness, which may be deeply affective.<sup>18</sup>

According to Damasio, the existence of subjectivity enables the brain/mind to know itself. It has knowledge of itself as an experiencer; it thus becomes the ‘I’ or knower. Damasio pays tribute to James’ account of the self, directly referring to James’ ideas in his own division of the self into the ‘me’ self and the ‘I’ self.<sup>19</sup> It does not matter whether or not the knowledge the ‘I’ has is accurate or inaccurate, that what is known is complete or incomplete. What matters is that there is something in the brain that is experiencing itself as itself, as a something. I made a similar point in Chapter 5 in relation to Descartes and the possibility of the evil demon. One can see that this runs counter to Dennett and Metzinger’s condition that selfhood requires accurate self-knowledge all the time and that lack of veridical knowledge is an argument against the existence of self.

---

<sup>18</sup> Panksepp and Northoff, 2009, p200.

<sup>19</sup> In Damasio’s account, these dual aspects represent the distinctions between his core self and the autobiographical self. A. Damasio, 2010, pp. 8-11.

Damasio posits three stages of self development. The first is the proto-self established by the existence of primordial (as opposed to primitive) feelings of embodiment. The second is the development of the core self which is grounded in the proto-self and which comes into being through the process of self-object awareness and the differentiation this entails. The third stage is the development of the autobiographical self which comes into being with the advent of language and the awareness of past and future engagement. The autobiographical self also contributes to continuity of self through providing a context and an identity that stretches over time.

Controversially, Damasio claims that consciousness requires a witness or knower in the system, some 'one' who sees images and knows/recognises that they are images of something. He claims there needs to be 'something extra' that enables the system to know that it is experiencing something and then to know what it is it is experiencing. These two acts create the feeling of knowing that there is something that knows. Damasio here is talking about the phenomenal element of consciousness, the state of being conscious of something. According to Damasio, there are (and have been) minds or mental processes without consciousness. Damasio claims that from an evolutionary perspective, there were mind processes before there were self processes. But it is only when we get subjectivity we get a knower, and it is only with the advent of a knower that we know minds exist. As he claims, he puts the witness back into the equation.<sup>20</sup> However, as he goes on to assert, this is not a non-physical witness. The mind/self/consciousness triad is firmly rooted in the body and produced by the processes of the body.<sup>21</sup> What Damasio appears to be intimating is that we ourselves are evidence of the knower because we know and we can reflect; we have this capacity. Without such a capacity, there could be no consciousness and no conscious deliberation.<sup>22</sup> This, he claims, is the core self. While this position seems to run counter to current (anti-Cartesian) thinking, consciousness itself is considered to be the last remaining 'hard

---

<sup>20</sup> Superficially, this seems like Parfit's 'further fact' which I address in a later section. Nevertheless, the ipseity does need to be explained and this is a problem for all accounts, not just Damasio's.

<sup>21</sup> Damasio, 2010, p.20.

<sup>22</sup> Damasio, 2010, p.17.



problem' in neuroscience.<sup>23</sup> If consciousness is viewed according to Dennett's model, then it can be reduced to just the functioning of specific parts of the brain that enable the system to respond appropriately. It becomes an easy problem. If it is the phenomenal experience of being (aware of) something, then it is not so easy to explain within a physicalist system. It becomes a hard problem. Damasio accepts that we are phenomenally conscious and that there really are experiences and states of being. As I stated in Chapter 4, the onus should be on those who claim otherwise to demonstrate that this is not the case. They need to show that there is nothing it is like to be or do anything, that we are really automatons with no inner world.

Damasio claims that this conscious discrimination of self as a something is the beginning of the self-process. As perception extends beyond the body to external objects, the sense of oneself as a being becomes more defined and refined. It starts to differentiate self from other and be aware of itself as a singularity. It is worth noting here that while Damasio appears to mirror Metzinger's position that the self is a process, there are important differences. First, Damasio is quite unequivocal that selves exist, thus his self-process constitutes a real self. In contrast, Metzinger uses the words 'self-process' to distance the self from objecthood and to support its illusory status.<sup>24</sup> Metzinger deliberately draws an ontological distinction between a self-as-process and a self-as-object, claiming the former has a different ontological status to the latter. Damasio's claim that the self is a process, however, is not a metaphysical claim. The self evolves and refines over time; it changes, therefore there is a process of development. For Damasio, both minds and selves are processual because they are not given complete at birth. They develop. Mind processes and self processes produce minds and selves, whether primitive or highly developed. But both exist.

---

<sup>23</sup> See special issues of *Journal of consciousness studies*, 'Explaining consciousness: the hard problem,' vols 2 & 3, 1995, 1996. But see also Glen Carruthers and Elizabeth Shier, *Dissolving the hard problem of consciousness*, for an alternative view. <http://consciousnessonline.com/2012/02/17/dissolving-the-hard-problem-of-consciousness/>.

<sup>24</sup> Damasio, 2010, p.8.

### 7.3 Damasio's self-model and the role of body maps

Damasio splits the cognitive domain into two distinct processes – mind processes and self processes. He then splits the self process into sub-types to ground them in more primitive systems and to create a bridge between low level and higher levels of the self. This means he posits what can appear to be three distinct self processes – the proto self, the core self and the autobiographical self. While this may look like he is positing multiple selves, this is not the case. In contradistinction to other models, these selves are not differentiated as conceptually or physically distinct.<sup>25</sup> For humans, each is a part of the one inevitable evolving or developing self. This is why Damasio talks about processes rather than entities. The fully-fledged self-conscious self is developing, in a similar way to Neisser's developmental model of the self.<sup>26</sup> In humans, the self-process reaches what Damasio calls the autobiographical stage.<sup>27</sup> This last stage requires recognition that there will be a future self and there was a past self, and it requires the capacity to express itself in language. This means fully-fledged human self-consciousness is unlikely in most animals we know. Within Damasio's framework, while the self can/may exist in a variety of different creatures in some form, it may never get beyond the core self stage or even the proto-self stage. Nevertheless, Damasio argues that consciousness only requires minimal subjectivity and this would be available to a wide range of diverse organisms, including those without brains. Thus, selfness would exist on a continuum, from minimal to maximal, depending on the individual organism's perceptual capacities and ability to discriminate self from other. By placing subjectivity as a basic or primary facet of sentient creatures, it is embedded in all levels of cognition. However, it is also available to life-forms other than humans or even animals.

---

<sup>25</sup> For example, Neisser posits five distinct selves. '[s]elf-knowledge is based on several different forms of information, so distinct that each one essentially establishes a different "self."' U. Neisser, 'Five kinds of self knowledge', *Philosophical Psychology*, Vol 1. Issue 1, 1988, pp. 35-59.

<sup>26</sup> At the same time, his is a developmental model that begins in infancy. By Neisser's own account, they are not experienced as distinct and, once developed, they all seem to be present to varying degrees in all activities.

<sup>27</sup> Damasio's autobiographic self is not the same as the narrative self discussed in Chapter 3. This autobiography represents the identity of a real self.

To explain how this happens, Damasio claims that the basis of all selfness is rooted in the body's capacity to form body maps. Body maps are 'special kinds of mental images of the body.'<sup>28</sup> In humans, the body-mapping structures are located in the primitive part of the upper brain stem. As discussed in Chapter 5, Metzinger uses the existence of body maps to demonstrate the highly representational nature of perception and to create a distance between perception and reality. In contrast, Damasio stresses that body maps are not simply representations of the body. He argues that the body and brain are bonded together in the process of constructing maps through 'resonance loops' that provide constant and dynamic feedback. This is a continual live cycle of constant ebb and flow that gives the brain a 'privileged and direct relationship to the body.'<sup>29</sup> In effect, there is no separation between perception and body at this level. This is why Damasio terms this the 'proto-self;' it is the first primitive stage of the self.

This is in sharp contrast to other accounts of the same phenomena. As we saw with both Dennett and Metzinger, there is no self at this level, just distributed information that needs to 'come together' at some other level first, in order for it to be represented as a perception of x. Only then can the system say 'this is x.' As I have argued elsewhere in this thesis, this position reduces consciousness to just the capacity to report on the contents of one's brain/mind. The self then gets posited much later. Damasio differs here because he claims that all perceptual information is experiential. Its interaction with the human system triggers a phenomenal reaction.<sup>30</sup> Thus, the first basic body maps become the first stages of self awareness because the body-mapping feels like something; it is accompanied by a feeling (a throbbing, a pulsing) of life, of existence. This is what Damasio terms the '*living body proper*.' This position takes phenomenal consciousness to be an existent fundamental property of sentient creatures. This phenomenal consciousness is necessarily self or subject-consciousness. Damasio wants to stress the

---

<sup>28</sup> Damasio, 2010, p.20.

<sup>29</sup> Damasio, 2010, p.21.

<sup>30</sup> Gerrans defines an affect program in 'Endnote 1, 'Delusional Misidentification as Subpersonal Disintegration,' *The Monist*, Vol.82, no.4, 1999, pp. 605-606. 'Affect' is the neurological substrate of emotions. Affect programs are associated with 'modularised neurological programs which have a specific range of elicitors and a distinctive phenomenology'. He claims there is evidence that perceptual modalities have an accompanying affect program which gives them their qualitative feel.

intimate relationship that exists at the organism level between its body and its awareness of pure process or of being alive.

Although this awareness is minimal at this stage in all creatures, the processes that create it are far from primitive. We see a similar complexity and refinement in Dennett's (far from minimal) biological self.<sup>31</sup> The first body maps and their accompanying phenomenology transmit information about the current condition of the body, its needs, the mechanisms for maintaining homeostasis, its intentions and goals for continuing survival and flourishing. They provide a very sophisticated, proto-conscious body of knowledge that contains a wealth of innate knowledge that can be used by the system.<sup>32</sup> When the body map changes to register interaction with an object, those changes enable the beginning of subjectivity as they acknowledge the body's existence in contradistinction to the object. That second-level awareness creates the core self in a mutually constitutive process.

Bermudez also places self-consciousness as something fundamental to systems with the right sort of architecture, present in both prelingual infants and many types of animals. Similar to Damasio, Bermudez claims this form of non-conceptual self-consciousness is rooted in our perceptual mechanisms at the bodily level of interaction with the environment. He lists four domains which could provide primitive forms of self-consciousness—perceptual experience, somatic proprioception, self-world dualism in spatial reasoning, and psychological interaction.<sup>33</sup> Bermudez refers to the work of J.J. Gibson who provides a detailed account of the way that visual information enters the visual system already delimited by self-structuring invariants. These invariants include limitations to the visual field produced by the narrowness of the visual field, the way one's own body occludes parts of the visual field, and the angles at which light hits off

---

<sup>31</sup> Referred to in Chapter 4. See Dennett, 1991.

<sup>32</sup> Damasio, 2010, pp. 35-36.

<sup>33</sup> Jose Luis Bermudez, *The Paradox of Self-consciousness*, MIT Press, Bradford, 1998, Chapters 5 & 6.

objects in one's visual field depending on one's position, height or movement, as well as the position, height or movement of the objects.<sup>34</sup>

Bermudez also explains how receptors in the hands and arms sensitive to skin stretch and joint movement or position provide information on the size, weight and shape of objects being touched as well as information about one's own bodily position, its strength and size in comparison to the objects. Again, the body adjusts and responds to salient environmental features in a double process of self-definition and environment definition, motivated by the relevance to the organism as a being with 'self' interests. Bermudez goes on to give a detailed account of the mechanisms of somatic proprioception—pressure receptors, temperature and friction receptors on the skin, joint receptors sensitive to stasis, muscle receptors, balance and posture receptors via the inner ear, nociceptors for bodily disturbances, receptors sensitive to gravity resistance and blood composition. Some of these systems provide information solely about the condition of the body, some provide relational information between the environment and the body, others provide information about both through touch, giving a clear (direct) indication of the limitations of the body and its position in relation to the environment. Somatic proprioception plays a major role in providing an internal perception of the body as well as its position in relation to itself and in relation to its environment. It is one way that the body both becomes aware of itself from the inside, and discerns between self and non-self, from the outside. If all these processes are afferent or have a distinct 'feel' then there would be strong grounds for claiming that the capacity to sense or perceive and a sense of self or subjectivity are simultaneous and fundamental. As stated earlier, they would be mutually constitutive. One would entail the other.

Taking a similar approach, Damasio spends a great deal of time discussing body maps because he wants to emphasise the close and continual contact between the body and the brain. He wants to emphasise the way they feed into each other. The continually updated and adjusted maps provide a constant image of the body. The images create sensations of feeling, feelings of embodiment that are present all the time one is awake. This means

---

<sup>34</sup> J.J. Gibson, 'Foundations of Ecological Optics: Part I' in *Reasons for Realism: Selected Essays of James J. Gibson*, (eds.) Edward Reed and Rebecca Jones, L. Erlbaum, Hillsdale, 1982.

that, when things are working properly, there is never a separation from body-mindedness. One is always body-minded, a possibility I raised against Strawson's purely mental phenomenon account of the self. However, in line with Strawson, Damasio claims this is not inferred but direct.

The continual mapping brings the body to mind all the time. This is the precursor of consciousness and the self. Once the awareness of body is presented as 'this body', we have proto consciousness and a proto self.

The living body is the central locus. Life regulation is the need and the motivation. Brain mapping is the enabler, the engine that transforms plain life regulation into minded regulation and, eventually, into consciously minded regulation.<sup>35</sup>

In his earlier work (*Descartes' Error*), Damasio referred to this as the constant background hum that provides the grounding for subjectivity and self. He wants to stress its constancy. He claims that this feeling of embodiment will provide a phenomenological sense of continuity between one moment and the next. While Metzinger also stresses the importance of body mapping or modelling, for Metzinger this process creates a distance between the constructed subject (the phenomenal self) and the experience. He acknowledges that it informs the phenomenal self model (PSM) but claims that what it creates is an illusion, a phenomenal self that is inferred not experienced. In contrast, Damasio sees the mapping creating a genuine link between moments of subjectivity. It provides a constant sense of sameness and continuity through change. He claims that this sensation grounds our experience of and belief in continuity, through the phenomenology of the body. This is in direct contrast to Strawson who argues that the self is a mental phenomenon, not a bodily one and, as a consequence, *self*-continuity is inferred not experienced. While Strawson acknowledges bodily sensations, he is sceptical of the EEE position of an embodied self because he claims

---

<sup>35</sup> Damasio, 2010, p107.

that the sensations from the body are not constant and vanish when we are engrossed.<sup>36</sup> Contra Damasio, he claims we experience our self as predominantly mental.

For Damasio, there is no disconnection between the self, subjectivity and bodily sensations, even in those moments of total immersion in an act. Damasio grounds the self and consciousness in the physical workings of the body which means the self is experienced simultaneously with the body. Thus, continuity of self would be experiential as I argued in Chapter 6, not just inferred. While the experience of continuity does not make something continuous, the lived body model does provide grounds for the continuity of bodily experience, hence self experience. If this is uninterrupted, it could count as genuine continuity. It would result in more evidence for continuity than we can directly provide for ordinary external objects that we grant continuity to. Their continuity is always inferred, whereas our self continuity would be experienced as well as inferred.

Damasio places a lot of emphasis on feeling and the generation of emotions. He claims that all images come with an afferent response that provides the ‘me’ connection. This includes objects that are part of me as well as objects that are not part of me but are related to me. These appear to represent the ‘scope of me’ (my words) or my domain. Recent research conducted by Cloutier and Macrae support the idea that ‘self’ involvement has more affective significance.<sup>37</sup> In their experiments, being consciously involved in an activity enhanced the memorability of information surrounding that activity, even when that involvement was trivial and the information indirect.

Damasio argues that the constant hum of life only becomes apparent when it goes awry, as with Cotard’s or anosognosia whose flattening effect I discussed in Chapter 5, or when it vanishes almost completely as with the loss of proprioception. In Sacks’ seminal case his patient, Christina, described herself as ‘pithed like a frog’ because of the loss of

---

<sup>36</sup> The EEE position is environmentally embedded, embodied, ecological, enactive. I discussed this in Chapter 6 in more detail.

<sup>37</sup> Jasmin Cloutier and C.N. Macrae ran a series of experiments that indicate that the act of choosing enhances memorability of information, no matter how trivial or indirect. ‘The feeling of choosing: self-involvement and the cognitive status of things past,’ *Consciousness and cognition*, 2008, Vol. 17, pp. 125-135.

internal position perception, a phenomenal sensation we take for granted and of which we would claim to be phenomenologically unaware.<sup>38</sup> For these reasons, Hamilton argues against the perceptual image or representation view of perception, particularly in relation to proprioception. He claims that proprioception is 1) not strictly a perception and 2) represents a case of direct knowledge.<sup>39</sup> He claims the ‘inner sense’ is not inferred but lived. It does seem to be the case that this type of perception is not ‘of’ something or some object. This may mean that it does not need to be re-presented in the cognitive system somewhere to enable a response, as would be the case for object awareness. The sensations or feelings are also not through the perceptual organs or systems such as visual, auditory or other sensory pathways. Heller-Roazen makes a distinction between proprioception and the inner sense, even though both provide a sense of ourselves from the inside that is normally non-cognitive. There does seem to be an aetiological distinction between the two in that a loss of proprioception is quite specific and the damage is localised. According to Sack’s accounts, while it is dramatically physically incapacitating, it does not appear to affect the patient’s sense of self or their physical and mental boundaries. Patients describe it as an ‘inner blindness;’ it results in a loss of internal sense of position that normally makes us aware of where our body is in space. This appears to be quite different from the effects of Cotard’s syndrome. According to the accounts of Cotard patients, Cotard’s affects their sense of self-embodiment and its boundaries. They claim to have a sense of their self expanding beyond its physical confines, along with a body that appears to be disintegrating. Yet this seems to have no effect on their physical capacity to control their body. Cotard’s is not a loss of proprioception but is more a loss of self definition or individuation.<sup>40</sup>

The distinction between Damasio, Dennett, Metzinger and even Strawson, comes to the fore when Damasio describes the beginnings of the self. He embeds phenomenal experience in every process, making consciousness and thereby subjectivity, essential to each perceptual process. He claims that the proto self is constituted by feelings. These

---

<sup>38</sup> Sacks, 1985, p. 30.

<sup>39</sup> A. Hamilton, ‘Proprioception as basic knowledge of the body,’ R. Woudenberg (ed), *The epistemology of belief*, Ontos Verlag, 2005.

<sup>40</sup> I discussed the effects of Cotard’s in some detail in Chapter 4.



feelings are the basis of all emotions. Panksepp and Northoff call this ‘affectivity’ and distinguish it from perceptual phenomenality.<sup>41</sup> Both contribute to subjective experience but the former is internally generated, while the latter is externally generated by interaction with the environment. Like Damasio, they claim affectivity or feeling is needed to generate value and meaning.<sup>42</sup> They claim this occurs through a process of self-relatedness which binds stimuli to self, enabling evaluation and motivation. One needs a reason to react to stimuli, whether internally or externally generated.

Affect systems appear to take that role by creating a sense of ‘mineness.’<sup>43</sup> Perception is thus active, not passive. In a related fashion, Stone and Young present a dual model of visual information processing, where the ventral route is for overt recognition of objective facts and the dorsal-limbic route provides affective information leading to an emotional response. They claim the latter is ‘implicated in the process whereby “relevance” is attached to an attended object’.<sup>44</sup> Capgras sufferers, for example, show overt recognition but no accompanying affect. As discussed in Chapter 5, neural imaging reveals the dorsal pathway is intact in prosopagnosics but not in Capgras (and Cotard) sufferers. Gerrans claims that ‘in the absence of affective processing, perception and cognition have no bodily consequences and thus are not ‘felt’ at the phenomenal level to belong to the agent.’<sup>45</sup> Damasio makes similar claims about patients with anosognosia.<sup>46</sup> They lack self interest. Thus, self-related (subjective) feelings and perception appear critical in making sense of the world and behaving appropriately.

---

<sup>41</sup> Panksepp and Northoff, 2009, p. 199.

<sup>42</sup> Panksepp and Northoff, 2009, p. 196.

<sup>43</sup> See Philip Gerrans, ‘Delusional Misidentification as Subpersonal Disintegration’ in *The Monist*, vol.82, no.4, 1999, and Tony Stone and Andrew Young, ‘Delusions and Brain Injury: The Philosophy and Psychology of Belief’ *Mind and Language*, Vol. 12, nos 3/4, Sept-Dec, 1997, p.595.

<sup>44</sup> Young refers here to Bauer’s model of visual processing. Bauer claims the dorsal pathway has ‘multiple functions encompassing automatic emotional responses to stimuli which have personal relevance.’ Andrew Young, Chap 10, ‘Recognition and Reality,’ *Face and Mind*, OUP, Oxford, 1998 p.265

<sup>45</sup> Gerrans, ‘Refining the Explanation of Cotard’s Delusion’ *Mind and Language*, vol. 15, no. 1, 2000 p. 119.

<sup>46</sup> I discussed this point in Chapter 4.

In terms of constancy, Damasio claims that interoception supplies the ‘invariance’ or stability that grounds the self. He claims this is critical because the self is a *singular* process and that singularity needs to be grounded.<sup>47</sup> It is grounded in the processes that change the least, which provide the most stable images. These come from the viscera and the internal milieu, the chemical broth, that only change slightly over time. He draws a distinction between the infinite variety of external stimuli and the resultant patterns they produce, and the infinite ‘sameness’ of the regulatory systems within the body and the maps they produce. This provides ‘an island of stability in a sea of motion.’<sup>48</sup> In addition, he claims that the mammalian brain stem is integral to producing the self. He also emphasises the ‘situated’ point of view provided by the sensory organs. These provide a location for the self and put it inside the body and centred in the head. He claims the proto-self, by design, is continuous; it is a continuous pulse of life, of feeling that provides the sense of sameness and stability. It is the grounding of a continuing self. While this is very much a phenomenal self, it is not an illusion as Dennett and Metzinger claim. Damasio says that in humans, it emerges from the neurological structure of the brain as a result of being a living creature. It is our way of being in the world and, contra Parfit, Dennett and Metzinger, there is no choice but to experience life from a position of selfness.<sup>49</sup>

Establishing the core self is the most critical aspect of Damasio’s theory. It places selfhood and consciousness at the intersection of body and world. The core self comes into being when there is self-object interaction. It provides a perspective and a sense of subjectivity. This perspective is not just a point of view as in ‘through these eyes.’ It is an awareness of looking at the world and engaging with the world through and from within one’s body. The core self is thus the advent of consciousness. Damasio claims

---

<sup>47</sup> Damasio, 2010, p193.

<sup>48</sup> Damasio, 2010, p200.

<sup>49</sup> Both Parfit and Metzinger intimate that lives would be better (less selfish) if we let go of the concept of self and moved towards the Buddhist ‘no-self’ philosophy. This is a simplistic view of Buddhism and is not true of many Buddhist doctrines. According to Bhikkhu, the ‘not-self’ doctrine in Buddhism is not a metaphysical position but a strategy for relieving suffering. After much effort, a condition of ‘not-self’ is attainable but it does not entail a loss of subjectivity or consciousness. Thanissaro Bhikkhu, *The not-self strategy*, 1993, Access to insight, June 5<sup>th</sup> 2010, [www.accesstoinsight.org/lib/authors/thanissaro/notself.html](http://www.accesstoinsight.org/lib/authors/thanissaro/notself.html).

that ‘the brain constructs consciousness by generating a self process within an awake mind.’<sup>50</sup> According to Damasio, although wakefulness and a mind are necessary, they are not sufficient for consciousness to emerge. You must have a self. One has to know that the thoughts, feelings and bodily actions are those of this entity. One is not just awake, as opposed to being asleep or comatose, but one is conscious of being conscious. This is the essence of subjectivity.

Damasio sets out four properties of the core self. There is a point of view or perspective; there is a feeling of being or existing; there is a sense of agency; and there is a sense of ownership of the contents of the mind. It needs to be noted that Damasio’s core self is not the same as Strawson’s minimal self, a point I raised in the last chapter. They are conceptually and empirically distinct. Neither can parallels be drawn between Damasio’s four essential properties and those that constitute Strawson’s minimal self. Strawson’s minimal self is created through a hypothetical stripping away of those properties that are inessential to constituting a sense of self. Damasio, on the other hand, is just describing the essential features of the self. One or more of these can be (and are) lost or damaged when fundamental parts of the brain stem are damaged. Although Metzinger also talks of consciousness and a point of view or perspectivalness, the PSM is not phenomenally conscious. Accessibility and agency are provided by his *phenomenal model of the intentionality relation* (PMIR).<sup>51</sup> He claims agency itself is an illusion because the information-processing system is non-conscious and action-mechanisms distributed.

In an interesting addendum, Damasio places biological value at the heart of all evolutionary processes. He claims that cells have in-built intentions and incentives that push them to maintain homeostasis and to seek flourishing. This is achieved through the existence of pleasure/pain/discomfort mechanisms which provide the incentive to avoid some things and seek out others because it makes the cell ‘feel better.’ He suggests that the advent of selves and consciousness brings awareness of intentions and value to the

---

<sup>50</sup> Damasio, 2010, p.180.

<sup>51</sup> As discussed in Chapter 5, Metzinger has two models – the PSM, which is the folk self and the PMIR, which is our self perspective. The PMIR produces action through ‘...an ongoing representational dynamics collapsing a phenomenal model of the practical intentionality relationship into a new transparent self-model’ (*Being no-one: The self model theory of subjectivity*, MIT Press, Mass, 2003, p. 563-4).

system. But it did not create them. They were already present in the system, embedded in each cell. Damasio argues that our own push to survive and flourish is grounded in the non-conscious intention of each cell. He then says we get a singular self/singular intention through having neurons. I do not have the capacity to comment on this view and it is one I have not found in other works. Nevertheless, it raises an interesting point about the source of intentions, interests and value. The organisation of an organic system is designed to perpetuate its own interests. Having a self-conscious or subjective system would be an advantage to perpetuating those interests. At the same time, an increasingly distinct and autonomous self could over-ride some of those interests, particularly if the social environment contains competing definitions of flourishing.

There is increasing evidence that the self or self-system plays an integral role in what we perceive, in what we remember and its significance, and in unifying action.<sup>52</sup> As already mentioned, if the conscious self is actively involved in an event, then one is more likely to impart significance, meaning or value to acts and objects such that one finds it easier to remember them. This is known as the self-referential effect (SRE). This effect has been shown to increase the memorability of information if it is self-referential rather than other referential. Gusnard also talks about the functional role the self plays.<sup>53</sup> Her studies using fMRI are starting to show that ‘having a self’ is critical in decision-making and behavioural control and that perception and behaviour may be tightly coupled together. The medial pre-frontal and parietal regions, especially on the right, have been associated with taking a first-person perspective in social and spatial situations.<sup>54</sup>

Damasio wants to stress the fact that cognition and consciousness are basic features of the human brain that occur through the early development of the mind and self. He argues that these come into being via feelings (pain, pleasure, discomfort) that are generated in the brain stem.<sup>55</sup> It is important for Damasio’s theory that he establish that

---

<sup>52</sup> Jasmin Cloutier and C.N. Macrae, ‘The feeling of choosing: self-involvement and the cognitive status of things past,’ *Consciousness and cognition*, 2008.

<sup>53</sup> D. A. Gusnard, Being a self: considerations from functional imaging, *Consciousness and cognition*, Vol 14, 2005, pp. 679-697 available [www.sciencedirect.com](http://www.sciencedirect.com).

<sup>54</sup> Gusnard, 2005, p.688.

<sup>55</sup> Damasio, 2010, p82-83.

the proto-self starts in the brain stem, rather than in the cerebral cortex. This part of the brain is the most primordial. If, as he claims, it is the seat of the self, then the self becomes an underlying feature of the architecture of the brain. This may be why amnesiacs can forget autobiographic details of the self but not memories associated with the early proto-self, like walking, talking, differentiating self from other, retaining a sense of self. He argues the same is true for consciousness. The advent of consciousness is aligned to the advent of self – proto-self and core self – and these he places in the (upper) brain stem, not in the cerebral cortex.<sup>56</sup> While he agrees that the cerebral cortex plays a major role in conscious processing, he claims it is not the seat of consciousness. According to Damasio, you need both, and possibly the thalamus.

#### **7.4 Problems with Damasio's self 'process'**

Damasio's account does not suffer from the same problems as Dennett's and Metzinger's in that he does not deny the phenomenology that is the basis for explaining consciousness and the self. The phenomenology is what drives his model. Feelings and sensations provide the impetus for self-generated action. Damasio also grounds his theory in neuroscience. Thus, his model unites our experience of being self-conscious selves with our developing knowledge of how the parts of the brain function. By placing the self and consciousness as underlying features of cognition, he enables the phenomenology to permeate all areas of information processing. This is a better fit with our experience and is supported by the changes to that self-experience when processes break down, as in the neuropathologies of the self discussed throughout this thesis.

Nevertheless, there are a couple of problem areas with Damasio's model. He claims that there has to be an internal 'witness' or knower in the system. This is his Jamesian 'I'. While he is careful to claim this is not the Cartesian non-physical witness, it is not clear what he means by this. As I indicated earlier, it could even be interpreted as a Parfitian 'further fact' as he does put it forward as 'in addition to' the experience of

---

<sup>56</sup> Here he differs from other neurologists.

subjectivity.<sup>57</sup> At the same time, he does not claim that the self is a thing or an object. Yet talk of a witness would imply an object-like entity more than a process. It is hard to see how a process can be a witness. It seems likely that the self-process produces an actual self that acts as a knower and a witness which has a *literal* point of view, generated by the sensory mechanisms and their location. Damasio is cautious about what he claims the self is, even while he is adamant that it exists. He does not enter directly into the ontological debate, possibly because he is not primarily a philosopher but a neuroscientist.

As mentioned earlier, I think the reason Damasio posits the self-as-process is indicated by the way he describes the advent of the core self. At one point he says that ‘...the core self states emerge in pulse-like fashion...’ but once a new stimulus occurs ‘...the whole cycle of conscious mind-making starts anew.’<sup>58</sup> In this passage, the core self appears to come into being when confronted by a stimulus that triggers the self-response, the identification of self as the subject of this experience. This ‘me’ moment appears to come at each new event. On the surface, this seems similar to Nozick’s self-defining moment of self-reflection or Strawson’s momentary conscious self.<sup>59</sup> As I argued in Chapter 6, this fleeting or ‘gappy’ self seems unable to support a robust sense of self and would not provide the continuity or unity of conscious experience that we claim to have, and which Damasio supports. It would not explain the continued existence of a self during non-REM sleep or in a sensory-deprivation tank. It also seems inconsistent with the robustness of the proto-self that lies at the basis of the core self and which enables its existence.

I think part of the problem is conceptual. One has to stop thinking of the core self as a discrete entity separate from the proto-self. As described above, most of the rudimentary work of establishing self-differentiation, stability, unity and continuity has been and is being carried out by the proto-self. This is constant. Like Dennett’s biological self, the

---

<sup>57</sup> Derek Parfit, ‘Personal Identity’ in *Reasons and Persons*, Oxford University Press, Oxford, 1986, p. 225.

<sup>58</sup> Damasio, 2010, p.209.

<sup>59</sup> Robert Nozick, *Philosophical Explanations*, Oxford, Clarendon Press, 1981.

proto-self is a sophisticated and complex entity; it is far from minimal. The transition from this pre-self to the core self is not fully explicated because there really is no sharp distinction between the selves. I think this is why Damasio talks about the self as a process. These are not discrete episodes or mechanisms. One stage or level flows into the other. The core self is built on and is part of the proto-self. There is, as he says, no dichotomy here. As soon as the proto-self encounters an object, the bodily adjustments trigger the core self, the awareness of change and that it is *me* changing. In his analysis, Damasio follows James' distinctions of 'me' and 'I'. The 'me' self is the basic self-as-object that knows which thing it is and what belongs to it or what is contained in it. The 'I' is the self-as-subject, the self as knower which comes on top of the 'me' self.<sup>60</sup> The 'me' self only has that sense of itself as existing, that it has boundaries and limitations. However, the 'I' self comes into being the moment (and for the moment Damasio refers to it as a 'pulse') there is an interaction with an object. This creates the sense of 'me' as 'I'. 'I' exist and am bounded in this body. I am conscious of being 'me'.

A good way of conceptualising this might be by drawing on Feinberg's description of a nested neural hierarchy. Feinberg has written extensively on 'where in the brain is the self?'<sup>61</sup> He, like Damasio, argues that the self is part of the neurology of the brain and that there are specific systems that work together to act as that self. While his model is different from Damasio's, he uses the concept of nested systems to demonstrate that each is a part of the other, rather like the rings of an onion, working from an inner core to an outer core.<sup>62</sup> He identifies three self-systems, the intereoself, exterioself and integrative self-systems. While these are hierarchical in that the intereoself system is about basic self maintenance and the integrative self is about the 'higher-order aspects' of the self, they do not operate in distinction from each other and aspects of each feed into aspects of the other. They are 'nested.' As Feinberg describes it,

---

<sup>60</sup> Damasio, 2010, p10.

<sup>61</sup> There have been two special editions of *Consciousness and Cognition* devoted to 'where in the brain is the self?', 'The Brain and Its Self,' 2005 and 'Brain and Self: Bridging the Gap,' 2011.

<sup>62</sup> TE Feinberg, 'The nested neural hierarchy of the self,' *Consciousness and Cognition*, Vol. 20, 2011, pp. 4–15.

In a nested hierarchy the lower levels of the hierarchy are physically combined or nested within higher levels to create increasingly complex wholes (also known as “holons”).<sup>63</sup>

According to Feinberg, this system results in a bottom up emergence model where the higher levels are composed of the lower levels and there is only weak top-down constraints. If one looks at Damasio’s model of the self, it could fit the nested hierarchy model more closely than a non-nested model. Thus, the core self in Damasio’s model is not a separate ‘holon’ but is constituted by the proto-self holon in a process of increasing complexity but where each stage is itself a ‘holon.’ The same would be true for the later autobiographical self.

For me, the most disappointing aspect of Damasio’s model of the self is his description of the autobiographical self. Damasio says that consciousness has enabled the development of rebellious selves, selves that can modify or go against biological derivatives such as homeostasis. Reflection means the beginning of freedom from automatic behaviour. I raised this point briefly when discussing the ‘intentions’ of cells. Dennett raised a similar point with the emergence of memes and their impact on the selfy-self. However, what Damasio then claims is inconsistent with his story to date. Damasio then says that establishing rules that reflect generalised well-being rather than individual well-being, was possible only through *story-telling*. He then says that ‘*implicit* storytelling has created our *selves*...’<sup>64</sup> Given what he writes about the proto-self and the core self, there seems little in Damasio’s account to justify such a strong statement. Stories and narratives need play no role in the existence of proto-and-core-selves, hence the positing of selves in other non-linguistic creatures. While the later autobiographical self can engage in *explicit* story telling, this does not entail *implicit* story-telling. Supposedly, selves were created by the act of perception, not by implicit story-telling. And again, who is constructing or telling the story? Even if it were the case that pre-lingual infants are inwardly talking to themselves in a proto-language or that non-lingual homonids have implicit narratives, which seems highly unlikely, Damasio doesn’t need it. His model works without it. To suddenly place narrative as the one and

---

<sup>63</sup> Feinberg, 2011, p. 11.

<sup>64</sup> Damasio, 2010, p.293.



only self-constructing mechanism undermines his whole model. It also introduces all the problems of authorship and ipseity addressed in Chapter 3, problems Damasio avoids with his proto-self.

In Damasio's model, selves can be sophisticated or rudimentary, depending on a creature's capacity to differentiate itself from its environment; the more rudimentary the perceptual capacity, the more rudimentary the sense of self or subjectivity. For example, a clam may only be capable of experiencing a sense of engulfing. It may be aware of light and warmth or movement. It will only have minimal subjectivity. A human, on the other hand, has a much more sophisticated sense of self, even without language, because it can perceive and discriminate so much more. Research findings reported by Rochat lend support to Damasio's claims about a basic or core self. He demonstrates that normally developing newborns already have subjective experience and minimal self-awareness.<sup>65</sup> As is clear from Damasio's own account, language alone does not create that self; it just increases its capacities for self dialogue and self reflection. It can talk to itself and, more importantly, it can talk to others. As I mentioned in Chapter 3, language enables the additional developmental effect of social interaction. It also allows others to engage with one's self, with all the effects on self this brings. It allows the presentation of a *persona* to the world. It presents a public face to what was an inner world. That doesn't mean it creates that self or its persona.

There is no denying the monumental impact that language has on our minds, our interactions, our capacities and, thus, our selves. However, Damasio doesn't need to place language as the creator of the self. It is the brain/body that develops those selves, that makes them possible, through body maps, interoception and exteroception. There is little to show that selves are dependant on language, even if the type of self is. Placing the advent of selves or subjectivity with phenomenal consciousness means language is neither a necessary nor a sufficient condition for selves to exist. So it is difficult to understand why Damasio does this, other than to conform to what has become the only story being told.

---

<sup>65</sup> P. Rochat, 'The self as phenotype,' *Consciousness and Cognition*, Vol. 20, 2011, pp. 109-119.

## **7.5 Consequences of Damasio's model: the self-as-system**

Despite the shortcomings mentioned above, Damasio's model actually demonstrates how a self could come into being through a developmental process from basic or proto-self to what Dennett refers to as our selfy self. I would probably label the stages of development as subject, self and then person, in line with the nature of the phenomenology attached to each stage and the developing social role of a linguistic self. This would mean that some creatures may be considered subjects and have subjectivity but they would not be selves. Given this framework, only humans would qualify as persons. However, it is important to retain the concept of embedded levels to prevent viewing the stages of development as distinct selves or discrete entities. Each stage is constituted by its previous stage. Depending on our activity, state of awareness and environmental demands, we may pass through or be in any of these stages. Diseases like Alzheimer's seems to be a good illustration of the plausibility of this type of developmental model, where it appears that the person is slowly stripped away, peeling back the layers to first a state of selfness and eventually back to a condition of simple subjecthood.

One could argue that a self-process is not the same as a self-object, that shifting from subjectivity to personhood does not support the idea of a singular diachronic self. However, one could equally argue that at any point in that process, there will be a subject of experience, there will be a point of view and there will be an agent capable of action and being taken as responsible for committing that action. This would appear to represent what we take to be a self. It is the nature of living organisms that they develop and mature, resulting in change. The self and its capacities are constantly developing and changing, both as part of the maturation/aging process and from exposure to social forces. If the process were to be impeded through illness, damage or some kind of neuropathology, such that the self-process ceased, it is likely that the 'person' in question would still act as though from a position of selfness. Where there is consciousness, there is a self, not just a self-process. Thus, primitive organisms, children and adults in various stages of health will all be potential subjects with non-existent, rudimentary or sophisticated selves.

There is increasing evidence that a self-like perspective is both shared by other mammals and plays a functional role in mediating and choosing appropriate behaviour. In experiments conducted to ascertain the existence of a SELF (Simple Ego-type Life Form in the brain), total neo-cortication in young mammals did not have an impact on coherent bodily reactions to stimuli or the mammals' ability to experience stimuli affectively.<sup>66</sup> The researchers went on to claim that,

...this ancient brain function allows SRP [self-related processing] to occur in homologous ways across all mammalian species and perhaps other vertebrates. ...the primordial coherence of this kind of core-SELF allows for epigenetic processing of external sensory-perceptual and cultural information within the extended, more cognitively oriented self in a self-referential manner that, in turn, enables organisms to selectively adapt to their respective environmental contexts in "self-projective" ways.<sup>67</sup>

This is a large claim. However, they show that some aspects of self activity are evident even when there is no cortical network. They put forward the idea that the presence of a SELF (identified above) is an early evolutionary development and that it is tied to emotional and affective responses that mediate interactions with the environment. As indicated by Gusnard and Coultier and MaCrae, having self dedicated structures enables creatures to react to stimuli in terms of their significance to their own needs or goals. This appears to support Damasio's grounding of the self and phenomenal consciousness in basic affective responses. If sensation and emotional content are critical for humans and animals to engage in self-preserving or self-serving behaviours, then it makes sense that these qualities are foundational to perception. Like Damasio, Northoff and Panksepp state, 'phenomenological experience of the core self is primary in affective processing rather than secondary.'<sup>68</sup> They also claim that the medial brainstem functions are critical for consciousness, homeostasis and emotional urges. What they postulate is a 'coherent' subcortical-cortical midline structure (SCMS) that is comprised of several neural systems whose purpose is to represent the self.<sup>69</sup> Again, in line with Damasio's position,

---

<sup>66</sup> Panksepp and Northoff., 'The trans-species core SELF: The emergence of active cultural and neuro-ecological agents through self-related processing within subcortical-cortical midline networks,' *Consciousness and Cognition*, vol 18, 2009 pp. 193-215 available [www.sciencedirect.com](http://www.sciencedirect.com).

<sup>67</sup> Panksepp and Northoff, 2009, p. 198.

<sup>68</sup> Northoff and Panksepp, 2008, p.259.

<sup>69</sup> Northoff and Panksepp, 2008, p. 259-264.

they claim that the higher mental and body-image self-processes are ‘intimately linked to the foundational neural “soil” of medial brain stem and medial frontal core-self functions.<sup>70</sup> One cannot separate the layers of self-functions into discrete entities or systems. They are intimately integrated, comprising a coherent system. They go on to state that,

it is important to conceptualize the self-integrative lower brain systems to have intimate relationships with the higher CMS structures, which is anatomically the case; thus yielding a highly integrated SCMS in normal organisms and providing a complex infrastructure for self-referential processing of external information.<sup>71</sup>

One could argue that the neurological data only indicates that self-like capacities or properties supervene on relevant brain states and that this is not evidence of a self or self-like structure with the sort of integrity we would need for objecthood. It does seem, however, that there is a growing body of research that is identifying very specific self mechanisms in the brain that do work together in such a way as to produce coherent self-phenomenology and self-related actions. Wherever these systems are located and however they are comprised, they could be a candidate for a physically-instantiated self. The research referred to here also places self and consciousness as primitive attributes of certain kinds of neuronal systems. Such findings add weight to the idea that having or being a subject, if not a self, is a natural developmental consequence for most (if not all) sentient systems. It is not just an outcome of the system but is basic to its functioning as a self-organised system, whose motivation is driven by its self-engagement with its world. By grounding the self and its phenomenology in fundamental structures in the brain, such as the brain stem, Damasio turns the self back into an object for scientific investigation; the self becomes knowable.

## **7.6 Establishing the self as object**

In the opening of this chapter, I said that I wanted to present a model of the self that seemed to deal fairly with our phenomenology without dismissing it and which was able to deal with the problems of the self. While there may be some discussion about the

---

<sup>70</sup> Panksepp and Northoff, 2009, p. 196.

<sup>71</sup> Panksepp and Northoff, 2009, p. 201.

exact neurological details of what subserves the self, there does appear to be increasing evidence that the self and consciousness are not higher order properties overlaying a non-conscious system. If we follow Damasio's model, consciousness and, therefore, the self or subject are primary features of human, if not all sentient systems. Consequently, Damasio's proto and core selves seem feasible candidates for object selfhood. Neurologists like Feinberg and Keenan or Panksepp and Northoff appear convinced that there is a well-integrated self-system in the brain. Vogely and Gallagher also intimate a similar finding, although they express some concern that the location of the self-system might be dispersed or distributed in the brain.<sup>72</sup> This need not be a problem if there are enough, robust connections between certain mechanisms to claim it acts as a singular system. The visual system would be an analogous case.

In Chapter I, I tried to demonstrate why the self is most often dismissed as non-existent and/or given no ontological reality. I argued that many of the reasons given can apply to many other everyday objects we take as existing in their own right. Metaphysics is a fraught area and there are a wide variety of disparate views about what can be said to exist. I attempted to show that establishing determinate identity conditions for the self is no more nor less a problem than it can be for all objects at some point and in certain circumscribed circumstances. It may be true that there is no fact of the matter, as Parfit claims. I argued that any object is susceptible to damage and destruction and, in this, the self is no different. In itself, this does not negate its existence. In fact, it is only concrete objects that are susceptible to change in this way. In Chapter 2, I set out what I believed to be the conditions a self would have to satisfy to be considered 1) an existent thing and 2) an object-like entity. These were:

- Unified boundedness or singularity – it must be identifiable as an entity in itself such that, if complex, it has an inherent unity of parts and that its boundaries can be delineated, enabling differentiation from other entities. It can individuate itself.
- Subjectivity – there must be a subject of experience such that the experiences are not just owned or identified as belonging to x but they are felt directly by x and as occurring within the parameters of x (selfness and consciousness).

---

<sup>72</sup> Kai Vogele and Shaun Gallagher, 'Self in the Brain,' *The Oxford Book of the Self*, OUP Gallagher S. ed, 2011, p.110.

- Synchronic and diachronic identity – for something to count as an entity it must satisfy certain identity conditions. Existence, sameness and persistence over time appear to be essential for this.
- Agency- for something to exist, it would have causal powers. For the self to exist, it should play a role in the cognitive network such that it has an impact on actions and is aware of its impact on actions

## 7.7 Concluding remarks

I have presented evidence throughout this thesis to demonstrate that there are specific mechanisms in the brain that ‘take care of’ self-like functions and which appear to play a constitutive role in our self-phenomenology. It has been my contention that these mechanisms work together to produce our sense of self. When these are malfunctioning, our sense of self becomes disturbed. I have also tried to demonstrate that selfness is a fundamental feature of the human cognising system. It is embedded in the act of perception such that we cannot but develop a sense of self. In line with Strawson’s constraints, I claim that those mechanisms that produce the sense of self can be considered a self, if there is enough unity or cohesion between the parts to warrant being classified as a singular object. Ultimately, it will be an empirical matter to determine whether or not such a self-system has the requisite amount of unity and integration to count as a single complex object. There is, however, enough evidence to show that (at least phenomenologically) a unified self is the norm and is the default position when things go wrong. It takes a lot to break that unity. Given that the self-system is identifiable and physically instantiated in the brain, it has a spatio-temporal location. It could, conceivably, be measured and weighed. Hence, it would classify as a concrete object.

I have claimed that it is foolish to deny that we are phenomenally conscious just because we may not be able to explain what that is or how it is possible. It is likely that such a property or state is a condition of being a living thing or a biological entity of a certain level of complexity. It is, however, outside the scope of this thesis to discuss consciousness directly. Nevertheless, one of my main arguments against the adequacy of the explanations of our phenomenology has been the denial of phenomenal consciousness. While personal and narrative identity theories address issues of identity,

they failed to acknowledge the sense of subjectivity that is at the heart of the problem of the self, its ipseity. We are self-conscious entities. While the subject of experience may be elusive, it makes no sense to posit just a collection of experiences without a perceiver of some kind. This perceiver is likely the product of the various sub-self systems working together to produce that experience and its individuation. Damasio's core self is constituted by the act of perception. The self-mechanisms set the boundaries of that self. Thus the subject is not just a capacity to react but is a state of being aware of itself as a subject and as a singular thing.

The synchronic and diachronic aspects of continuing identity are the hardest for the self to satisfy. If the self is constituted by a system of mechanisms or processes in the brain, then it is tied to its physical instantiation within that brain. Where does that leave the changing psychology of the self or its personal identity, the things that make it a unique self? As I have tried to demonstrate throughout this thesis, this is a problem that is not unique to the self. It impacts on all changing objects. At the same time, I have attempted to separate qualitative identity from quantitative identity. Who you are as a particular instantiation of a self will be determined by your history, social relationships and events, and your developing character. These are part of you and your neurology. Who you are as a self will be determined by your unique instantiation as a particular individuated consciousness which, according to Gibson, is individuated by its particular embodiment that produces its perspective or point of view. Strawson argued that only the singular moment of subjectivity exists. Each subsequent moment of subjectivity is a different subjectivity or self. I argued that there is enough continuity in the system for these moments to count as continuous moments, given that perception at all levels is a conscious phenomena. If the body as a perceiving thing continues then awareness of self continues. It is not as gappy as we think. Even so, the self-system continues as a physical entity even if our sense of self is discontinuous.

Strawson argued that agency is not an essential feature of a self because it was not essential to having a sense of self. It is true that some people have no capacity for action, hence agency, because of extreme physical debility. One can also lose control of one's body through use of drugs or anaesthetics. Nevertheless, it seems part of our

phenomenology to have a sense of control over our bodies in normal circumstances and to believe that we (our self) can have a causal impact on the world around us. Narrative accounts leave decision-making and action to something other than the self. Dennett gave this role to memes. Yet, there does appear to be evidence that humans can engage in actions and decision-making that seem to involve a thinker and doer. While this may not be essential for subjectivity, it is for a sense of agency. Given the limitations of this thesis, I have not been able to address agency in any detail. Yet, if it is the case that selves do exist, then they will have a causal impact in the world. I claim that denying self-agency ignores an important part of the reality of our social world. Research indicates that self-involvement has an impact on how the human cognitive system operates. As Nahmias demonstrates in his study, ‘conscious intentions are [normally] causally implicated in actions.’<sup>73</sup> Self-conscious intentions can change the way we act, even when such changes requires extreme effort. Giving up an addiction is a case in point. If we want something enough, we can do amazing things.

I have attempted to show that the grounds used to dismiss the self are based on a rejection of the Cartesian self and an unwillingness to conceive of the self as other than an expression of our personal identity. My self may reflect a particular persona but the having of a self or experiencing life from a position of selfness is not just reducible to that identity. The self encompasses much more than its personality and autobiographical details. To claim otherwise is to ignore our sense of subjectivity and our sense of agency, both which could exist without our respective identities. As stated above, whether the self-system I have proposed counts as a concrete object or not will be determined by future research. Nevertheless, I believe there are strong grounds to consider the self to be constituted by such a system, given its role in producing our self phenomenology. Contrary to Parfit’s claims, I doubt we have a choice but to experience ourselves from a position of selfness; it is likely a biological imperative and one that is essential for our survival. The difficulties inherent in coming from a position of non-self

---

<sup>73</sup> Nahmias, E. ‘Agency, authorship and illusion’ in *Consciousness and Cognition*, vol. 14, 2005, p. 780.



seem to support this.<sup>74</sup> As such, I believe there are sufficient grounds to claim that the self does exist and that it could be considered object-like. It is plausible to consider it as real as a toaster or as real as the visual system. How real one considers that will depend on one's ontology.

---

<sup>74</sup> Parfit (and Metzinger) both suggest that the Buddhist approach of 'no-self' would be better and make for a less 'self-ish' world. In the 'Not-self Strategy,' Thanissaro Bhikkhu disputes that Buddha posits a no-self ontology. He claims that the idea of moving away from 'self' in Buddhism is a life strategy, not a metaphysical position. He also emphasizes the extreme difficulty of reaching such a position. Thanissaro Bhikkhu: (Geoffrey DeGraff)", edited by Access to Insight. *Access to Insight (Legacy Edition, version ati-legacy-2013.12.28.08)*, 1994, accessed June, 2010. <http://www.accesstoinsight.org/lib/authors/thanissaro/index.html#notselfstrategy>

## BIBLIOGRAPHY

Aaron, R. I., *The Theory of Universals* (second edition), Oxford University Press, London, 1967.

Alakkalkunnel F and Kanzian C., *Strawson's Concept of Person – A Critical Discussion*, 2002, accessed April, 2013. <http://sammelpunkt.philo.at:8080/1349/1/alakka.pdf>

Allison R.B., 'On discovering multiplicity' in *Svensk Tidiskrift fur Hypnos*, 1978, vol 2, pp. 4-8.

Allison R., 'Difficulties diagnosing the multiple personality syndrome in a death penalty case,' *The International Journal of Clinical and Experimental Hypnosis*, vol. 32, No. 2, 1984, pp102-117.

American Diagnostic Statistic Manual, *DSM IV*, <http://www.dnalc.org/view/899-DSM-IV-Criteria-for-Schizophrenia.html>.

Anderson, P., 'A Question of Personal Identity,' *Personalist Forum*, 1992, 8 (1), pp. 55-68.

Armstrong D., 'How do particulars stand to universals?' in *Oxford Studies in Metaphysics: vol. 1*, (ed) Dean Zimmerman, OUP, 2004.

Armstrong, D., *Universals: an opinionated introduction*, Westview Press, Colorado, 1989.

Ayer, A. J., *Hume*, Oxford Press, London, 1980.

Baron-Cohen, S., 'Autism – 'autos': literally, a total focus on the self?' in *The Lost Self: Pathologies of the Brain and Identity*, eds. T E Feinberg and J P Keenan, 2005, Oxford University Press, London.

Baars, B. J., *Cognitive Theories of Consciousness*, Cambridge University Press, London, 1988.

Beckermann A, Flohr H, Kim J (eds), *Emergence or reduction? Essays on the Prospects of Nonreductive Physicalism*, Berlin, De Gruyter, 1992.

Bermudez, J. L., 'Ecological perception and the notion of a nonconceptual point of view' in J.L Bermudez, A. Marcel, and N. Eilan (eds) *The body and the self*, MIT Press, Massachusetts, 1998.

Bermudez, J. L., *The Paradox of Self-Consciousness*, MIT Press, Bradford, 1998.

Bermudez, J. L., 'Precis of The Paradox of Self-Consciousness' *Psycoloquy*, 1999, 10.35.

Bermudez, J.L, Marcel A., and Eilan N. (eds) 'Introduction,' in *The body and the self*, MIT Press, Massachusetts, 1998.

Bhikkhu T., 'The Not-self Strategy,' (ed) Geoffrey DeGraff, *Access to Insight*, 1994, <http://www.accesstoinight.org/lib/authors/thanissaro/index.html#notselfstrategy>.

Bigelow, J., 'The Ontology of Scientific Realism', *The Monist*, Jan, 1994.

Blackburn, S., 'Has Kant Refuted Parfit?' *Reading Parfit*, Jonathan Dancy (ed), Blackwell Publishers, Oxford, 1997.

Blackmore, S *The Meme Machine*, Oxford University Press, 1998, p. 230.

Blakemore, C. and Greenfield, S., *Mindwaves*, Basil Blackwell, Oxford, 1987.

Block Ned, 'Troubles With Functionalism,' in *Readings in Philosophy of Psychology*, Ned Block (ed), Cambridge, Harvard University Press, 1980a. [http://www.mcps.umn.edu/philosophy/9\\_12Block.pdf](http://www.mcps.umn.edu/philosophy/9_12Block.pdf).

Block, Ned, 'Begging the Question Against Phenomenal Consciousness' *Behavioral and Brain Sciences*, 1992, 15:2, pp. 205-206.

Block, Ned, 'Review of *Consciousness Explained*', *Journal of Philosophy*, 1993, XC, no. 4, pp. 181-192.

Bowman E. and Coons P., 'The differential diagnosis of epilepsy, pseudoseizures, dissociative identity disorder, and dissociative disorder not otherwise specified,' *Bulletin of the Menninger Clinic*, vol 64, no. 2, 2000, pp. 164-180.

Boyer P, Robbins P and Jack A, 'Varieties of self-systems worth having,' in *Consciousness and Cognition*, vol 14, 2005, pp 647-660.

Butchvarov P., 'Metaphysics' in *The Cambridge Dictionary of Philosophy*, 2<sup>nd</sup> ed., ed. R. Audi, Cambridge University Press, U.K., 1999, pp.563-566.

Carroll R.T., 'Multiple Personality Disorder,' <http://skepdic.com/mpd.html>.

Carruthers, Peter, 'Review of 'The Paradox of Self-Consciousness' by Jose Luis Bermudez, *British Journal for the Philosophy of Science*, 1999, 50, pp. 483-486.

Carruthers G. and Shier E., *Dissolving the hard problem of consciousness*, <http://consciousnessonline.com/2012/02/17/dissolving-the-hard-problem-of-consciousness/>, accessed November, 2013.

Carruthers G., 'A Model of the Synchronic Self' in *Consciousness and Cognition*, vol 16, pp 533-550, 2007.

Cassam Quassim, 'Introspection and bodily self ascription' in *The Body and the Self*, ed Bermudez, Marcel and Eilan, MIT Press, 1995.

Bowman, E and Coons, P., 'The differential diagnosis of epilepsy, pseudoseizures, dissociative identity disorder, and dissociative disorder not otherwise specified,' *Bulletin of Menninger Clinic*, 2000, vol 64, no. 2, pp. 164-180.

Chakravartty Anjan, *A Metaphysics for Scientific Realism*, Cambridge University Press, Cambridge, 2007.

Chalmers Alan, *What is this thing called science?* 3<sup>rd</sup> edition, UQP, St. Lucia, 2006.

Chalmers, David, *The Conscious Mind*, Oxford University Press, New York, 1996.

Chalmers D., 'Naturalistic dualism' in M. Velmans and S. Schneider (eds) *The Blackwell Companion to Consciousness*, Oxford: Blackwell, 2007, pp 359-368.

Chalmers David, 'Ontological anti-realism' in *Metametaphysics: New Essays on the Foundations of Ontology*, (eds) David Chalmers, David Manley, and Ryan Wasserman, OUP, New York, 2009.

Chalmers D., Manley D, and Wasserman R., (eds) *Metametaphysics: New Essays on the Foundations of Ontology*, OUP, New York, 2009.

Chi M., Slotta J. and Leeuw N., 'From things to processes: a theory of conceptual change for learning science concepts,' in *Learning and Instruction*, vol. 4, pp 27-43, 1994.

Chomsky, N, *Language and Mind*, Harcourt Brace Jovanovich, New York, 1972.

Churchland Patricia, *Neurophilosophy: towards a unified science of mind-brain*, MIT Press, Mass., 1986.

Churchland, P. M., *Matter and Consciousness*, (revised edition), MIT Press, Cambridge, 1988.

Cloutier, J and Macrae, C.N *Consciousness and cognition*, 2008, Vol. 17, pp. 125-135.

Coburn, R.C., 'Personal Identity Revisited', *Canadian Journal of Philosophy*, 1985, Vol 15, No 3, Sept., pp. 379-404.

Coltheart M., Langdon R., and McKay R., 'Schizophrenia and Monothematic Delusions,' *Schizophrenia Bulletin*, 2007, vol. 33 no. 3 pp. 642-647.

- Coulter, J., *The Social Construction of Mind: studies in ethnomethodology and linguistic philosophy*, Totowa N.J., Rowman Littlefield, 1979.
- Couvalis, G., *The Philosophy of Science: Science and Objectivity*, London, Sage Publications, 1997.
- Dahlbom, Bo (ed.), *Dennett and His Critics: Demystifying Mind*, Oxford, Blackwell, 1993.
- Damasio A. R., *Self Comes to Mind*, Pantheon, New York, 2010.
- Damasio, A. R. 'Mental self: The person within,' *Nature*, vol. 423, no. 227, 2003.
- Damasio, A. R., 'Feelings of emotion and the self,' *Annals New York Academy Sciences 1001*, 2003, pp. 253-261.
- Damasio, A. R., *Descartes' Error: Emotion, Reason and the Human Brain*, London, Papermac, 1996.
- Damasio, A. R., 'The Selfless Consciousness' *Behavioral and Brain Sciences*, 1992, 15:2, pp. 205-206.
- Dancy, J. (ed.), *Reading Parfit*, Blackwell Publishers, Oxford, 1997.
- Daprati E., Franck N., Georgieff N, Proust J., Pacherie E., Delany J., Jeannerod M., 'Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients,' *Cognition* 65, 1997.
- Davidson, D., *Inquiries into Meaning and Truth*, Oxford University Press, U.S.A, 1990.
- Davies, Bronwyn and Harre, Rom, 'Positioning: The Discursive Production of Selves', URL:<http://www.massey.ac.nz/~Alock/position>, 1999.
- Debruyne H, Portzky M, Van den Eynde F., and Audenaert K, 'Cotard's syndrome: a review' *Current Psychiatry Reports*, vol.11(3), 2009, pp 197-202. <http://www.ncbi.nlm.nih.gov/pubmed/19470281>.
- Dennett and Kinsbourne, 'Time and the Observer: the where and when of consciousness in the brain', *Behavioral and Brain Sciences*, 1992, 15:2, June, pp. 183-200.
- Dennett, D. C., 'Conditions of Personhood' *The Identities of Persons*, Amelie Rorty (ed.), University of California Press, California, 1976.
- Dennett, D. C., *The Intentional Stance*, MIT Press, Cambridge, Mass., 1987.
- Dennett, D. C., *Consciousness Explained*, Little, Brown, Boston, 1991.

- Dennett, D. C., 'Real Patterns' *The Journal of Philosophy*, 1991, 87, pp. 27-51.
- Dennett, D. C., 'The Self as a Center of Narrative Gravity', F. Kesses, P. Cole and D. Johnson (eds.), *Self and Consciousness: Multiple Perspectives*, Erlbaum, Hillsdale, N.J., 1992.
- Dennett, D. C., 'Precis of Consciousness Explained', *Philosophy and Phenomenological Research*, 1993, LIII, No. 4, pp. 889-892.
- Dennett, D. C., 'The Message is: There is no Medium', *Philosophy and Phenomenological Research*, 1993, LIII, No. 4, pp. 919-931.
- Descartes, R., *The Philosophical Works of Descartes*, trans. Elizabeth S. Haldane and G. R. T. Ross, Cambridge University Press, Cambridge, 1973.
- Dissociative Fugue Symptoms, *Psych Central*. Retrieved October 9, 2013, <http://psychcentral.com/disorders/dissociative-fugue-symptoms/>.
- Duggal, H.S., Jagadheesan K, and Haque Nizamie S., 'Biological basis and staging of Cotard's Syndrome,' *Eur Psychiatry*, 2002, vol.17, pp. 108-9.
- Egege Y.S., *Dennett and the self: is the self a social construction?* MA (Phil) Thesis, Flinders University, 2002.
- Elman, J.L, Elizabeth Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi and Kim Plunkett; *Rethinking Innateness: A connectionist perspective on development* Massachusetts, MIT Press, 1998.
- Everett D., *Language*, Profile Books, 2012.
- Feinberg T E., Eaton LA., Roane D M and Giacino J T. 'Multiple fregoli delusions after traumatic brain injury,' *Cortex*, 1999, vol. 35, pp 373-387.
- Feinberg T E and Keenan J P., *The Lost Self: Pathologies of the Brain and Identity*, Oxford University Press, London, 2005.
- Feinberg T. E. and Keenan J. P., 'Where in the brain is the self?' in *Consciousness and Cognition*, vol 14, 2005, pp. 661–678.
- Feinberg T E., 'Brain and Self: Bridging the Gap.' *Consciousness and Cognition*, vol 20, 2011, pp 2–3.
- Feinberg T E., 'The nested neural hierarchy and the self,' *Consciousness and Cognition*, vol 20, 2011, pp 4–15.
- Feinberg T E., 'Neuropathologies of the self: Clinical and anatomical features,' *Consciousness and Cognition*, vol 20, 2011, pp 75- 81.

Fellows, Roger and O'Hear, Anthony, 'Consciousness Avoided' - Symposium: Daniel C. Dennett's Consciousness Explained, *Inquiry*, 1992, 36, pp. 73-91.

Feyerabend, Paul, *Realism, Rationalism and Scientific Method: vol. 1*, Cambridge University Press, N. Y., 1981.

Fichte J. G., *The Science of Knowledge 1794-5*, trans. Peter Heath and John Lachs, CUP, 1982.

Flage, Daniel E., *David Hume's Theory of Mind*, Routledge, London, 1990.

Flanagan, Owen, *The Science of the Mind* (2<sup>nd</sup> edition), MIT Press, Cambridge, Mass., 1991.

Foster, John, 'Dennett's Rejection of Dualism' - Symposium: Daniel C. Dennett's Consciousness Explained, *Inquiry*, 1992, 36, pp. 17-31.

Fotopoulou, A., Jenkinson, P.M., Tsakiris, M., Haggard, P., Rudd, A. & Kopelman, M.D, 'Mirror-view reverses somatoparaphrenia: Dissociation between first- and third-person perspectives on body ownership.' *Neuropsychologia*, 49, 2011.

Freud S, *Civilization and its discontents*. Standard edition, Vol. 21, London, Hogarth Press, 1930.

Frith C., 'The self in action: Lessons from delusions of control,' in *Consciousness and cognition*, vol 114, 2005.

Frith C., 'Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action,' *Psychological Medicine*, vol 19, no. 2, 1989.

Frith C., 'The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action,' *Psychological Medicine*, vol 17, no. 3, 1987.

Gallagher S. (ed), *The Oxford Book of the Self*, OUP, 2011.

Gallagher S., 'Self reference and schizophrenia: a cognitive model of immunity to error through misidentification,' in *Exploring the Self: philosophical and psychological perspectives on self experience*, ed. Dan Zahavi, John Benjamins Publishing Company, The Netherlands, 2000.

Gallagher S., 'Philosophical conceptions of the self: implications for cognitive science' in *Trends in Cognitive Sciences*, vol 4, no. 1, 2000, pp 14-21.

Gallagher S and Shear J (eds), *Models of the self*, Imprint Academic, UK, 1999.

Gerrans, Philip, 'Cognitive Disintegration and Delusional Misidentification', *Psychology*, 1998.

Gerrans, Philip, 'Delusional Misidentification as Subpersonal Disintegration', *The Monist*, 1999, 82, No. 4, pp. 590-608.

Gerrans, Philip, 'Redefining the Explanation of Cotard's Delusion', *Mind and Language*, 2000, vol. 15, No. 1, pp.111-122.

Gibson. J.J, 'Foundations of Ecological Optics: Part I' in *Reasons for Realism: Selected Essays of James J. Gibson*, (eds.) Edward Reed and Rebecca Jones, L. Erlbaum, Hillsdale, 1982.

Gillett, Grant, 'Brain Bisection and Personal Identity', *Mind*, 1986, 95, pp.224-229.

Gillett, Grant, 'The Subject of Experience', *Logos*, 1990, 11, pp. 93-109.

Gillig, P., 'Dissociative Identity Disorder: A Controversial Diagnosis,' *Psychiatry*, 2009, vol. 6, no. 3, pp 24-28.

Gleaves, D. H., 'The Sociocognitive Model of Dissociative Identity Disorder: A re-examination of the evidence.' *Psychological Bulletin*, 1996. Vol. 120, No. 1, pp. 42-59.

Gusnard D. A., 'Being a self: considerations from functional imaging,' in *Consciousness and Cognition*, vol 14, pp. 679-697, 2005, available [www.sciencedirect.com](http://www.sciencedirect.com)

Hacking I, *Rewriting the Soul: multiple personalities and the sciences of memory*, Princeton University Press, New Jersey, 1995.

Hamilton, A, 'Proprioception as basic knowledge of the body,' R. Woudenberg (ed), *The epistemology of belief*, Ontos Verlag, 2005.

Hannan, B., *Subjectivity and Reduction*, Westview Press, USA, 1994.

Hardcastle V. (ed), 'Explaining consciousness: the hard problem,' *Journal of consciousness studies*, vols 2 & 3, 1995, 1996.

Hardcastle, V., 'A Self Divided: a review of *Self and Consciousness: Multiple Perspectives*' in *Psyche*, vol 2, no.1, 1995.

Hart W.D, *The engines of the soul*, Cambridge University Press, 1988.

Harter S., *The construction of selves: a developmental perspective*, Barnes and Noble, 1999.



Haugeland, John, *Artificial Intelligence: The Very Idea*, Bradford Book, MIT Press, Mass, 1987.

Haugeland, John, 'Patterns and Being,' in *Dennett and His Critics Demystifying Mind*, Bo Dahlbom (ed.), Blackwell, Oxford, 1993.

Heffner, Gary J, 'Cotard Syndrome Website,' 1999, accessed May, 2008/2011. <https://sites.google.com/site/autismhome/Home/miscellaneous/cotard-syndrome>

Heil John, *from an Ontological Point of View*, Clarendon Press, Oxford, 2003.

Heinrichs W, *In search of madness*, OUP, New York, 2001.

Heller-Roazen D, *The Inner Touch: archaeology of a sensation*, Zone Books, New York, 2007.

Hopkins Gerard Manley, *The Poems of Gerard Manley Hopkins*. Ed. Robert Bridges. London: OUP, 1930.

Hume, David, *A Treatise of Human Nature*, 2nd Edition, index by L. A. Selby-Bigge, Oxford University Press, 1985.

Hutchins, Robert M., *Locke, Berkeley, Hume*, William Benton (ed.), Encyclopaedia Britannica, Inc, 1952.

Hutto D. D., 'Folk Psychology as Narrative Practice,' *Journal of Consciousness Studies*, 2009, vol. 16, No. 6–8, pp. 9–39.

Hutto D. D. (ed), *Narrative and Understanding Persons*, Cambridge, Cambridge University Press, 2007.

James W, 'The Principles of Psychology (1890),' *Classics in the History of Psychology*, accessed April 2011, <http://psychclassics.yorku.ca/James/Principles/prin10.htm>.

Jarvilehto T, 'Efferent Influences on Receptors in Knowledge Formation', *Psychology*, 1998, 9, Issue 41.

Johnston, M., 'Human Beings', *The Journal of Philosophy*, 1987, LXXXIV, No. 2, pp. 59-83.

Johnston, M., 'Human Concerns With Superlative Selves', *Reading Parfit*, Jonathan Dancy (ed.), Blackwell Publishers, Oxford, 1997.

Kanaris, Jim, 'Engaged Agency and the Notion of the Subject', *Method: Journal of Lonergan Studies*, 1996, 14, No.2, pp.183-200.

Keleman, J., 'The Dynamics of Personality. An East European View', *Acta Neurochirurgica*, 1988, Suppl. 44, 33, pp. 33-38.

Kenny A., *The Self*, Marquette, Marquette University Press, 1998.

Kesses F., Cole P. and Johnson D. (eds.), *Self and Consciousness: Multiple Perspectives*, Erlbaum, Hillsdale, N.J., 1992.

Kierkegaard Søren, *The Concept of Anxiety* trans. Reidar Thomte, Princeton, NJ, Princeton University Press, 1980.

Kim Jaegwon, *Physicalism, or something near enough*, Princeton University Press, Princeton, 2005.

Kirscher, T and David, A (eds.), *The Self in Neuroscience and Psychiatry*, eds., CUP, 2003.

Kluft RP. Dealing with alters: a pragmatic clinical perspective *Psychiatr Clin North Am* 2006; 29 (1):281–304.

Kristinsson S., 'Authenticity, identity, and fidelity to self,' *Homage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz*. Eds. T. Rønnow-Rasmussen, B. Petersson, J. Josefsson & D. Egonsson, 2007, [www.fil.lu.se/HomageaWlodek](http://www.fil.lu.se/HomageaWlodek).

LaCaze, Marguerite, 'Splitting the Difference: Between Young and Fraser on Identity Politics' *Feminist Alliances*, Pluto Press, London, 2001.

Lane, A, 'The Narrative Self-Constitution View: Why Marya Schechtman Cannot Require it for Personhood,' *Macalester Journal of Philosophy*, Vol. 20, Issue 1, 2012

Leiber, Justin, 'Re(ad) Me; Re(ad) Myself' *Proceedings of 13th Annual Conference on Film and Literature*, Florida State University, 1988.

Legerstee M, 'Mental and bodily awareness in infancy: consciousness of self-existence,' in *Models of the Self*, editors S Gallagher and J Shear, Imprint Academic, U.K., 1999.

Libet B., Gleason C. A., Wright E. W., and Pearl D. K., 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act,' *Brain*, vol. 106, pp 623-642, 1983.

Lilienfeld S., Lynn S., Kirsch I., Chaves J., Sarbin T., Ganaway G. and Powell R., 'Dissociative Identity Disorder and the Sociocognitive Model: Recalling the Lessons of the Past,' *Psychological Bulletin*, 1999, Vol. 125, No. 5, pp.507-523.

Locke, John, Bk II, Chap. Xxvii, s 8 in *An Essay Concerning Human Understanding*.

Lockwood, M., 'Dennett's Mind' - Symposium: Daniel C. Dennett's Consciousness Explained, *Inquiry*, 1992, 36, pp. 59-72.

Lowe E. J., *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*, Clarendon Press, Oxford, 2006.

Lowe E. J., *The Possibility of Metaphysics: Substance, Identity and Time*, Clarendon Press, Oxford, 1998, p.1-27.

Lübcke P., 'A Semantic Interpretation of Husserl's EpochÉ,' *Synthese*, 1999, Vol. 118, No.1, pp 1-12.

Mackie, J. L., *Problems From Locke*, Clarendon Press, London, 1976.

Manley D., 'Introduction,' *Metametaphysics: New Essays on the Foundations of Ontology*, (eds) David Chalmers, David Manley, and Ryan Wasserman, OUP, New York, 2009.

Maxwell, N, 'Three philosophical problems about consciousness and their possible resolution,' in *Open Journal of Philosophy*, Vol. 1, no. 1, 2011, pp. 1-10. DOI:10.4236/ojpp.2011.11001

McDowell, J., 'Reductionism and the First Person', *Reading Parfit*, Jonathan Dancy (ed.), Blackwell Publishers, Oxford, 1997.

Merleau-Ponty, Maurice, *Phenomenology of Perception*, Routledge and Kegan Paul, London, 1962.

Metzinger T, *Being no-one: The self model theory of subjectivity*, MIT Press, Mass, 2003.

Metzinger T. and Gallese V., 'The emergence of a shared action ontology: Building blocks for a theory', *Consciousness and Cognition*, vol.12, 2003, pp 549–571.

Morito, Bruce, 'Fundamental Ontology and Personal Identity: A Critique of Albert Shalom's View of Personhood', *Review of Metaphysics*, 1989, 42, June, pp. 797-815.

Multiple personality disorder [dissociative identity disorder], <http://skepdic.com/mpd.html> accessed September, 2006.

Murphy G., 'Why there is no such thing as introspection,' paper presented to philosophy research seminar series, Adelaide University, June 2009.

Nagel, T., 'Brain Bisection and the Unity of Consciousness', *Synthese*, 1971, XXII, No. 3/4, May, pp.396-413.

Nagel T., 'What is it like to be a bat?' in *Philosophical Review*, 1974, pp. 435-50.

Nahmias, E. 'Agency, authorship and illusion' in *Consciousness and Cognition*, vol. 14, 2005, pp 771-785.

Neisser U., 'Five Kinds of Self-Knowledge,' *Philosophical Psychology*, vol 4, no1, pp 35-59. 1991.

Nicolson, Stewart, 'Illusions of Consciousness', *Dialogue*, 1995, XXXIV, pp. 769-775.

Noonan, Harold, *Personal Identity*, Routledge, London, 1989.

Northoff G. and Panksepp J., The trans-species concept of self and the subcortical-cortical midline system, *Elsevier*, 2008, doi:10.1016/j.tics.2008.04.007

Northoff G, Qin P, and Feinberg T., 'Brain imaging of the self – Conceptual, anatomical and methodological issues,' *Consciousness and Cognition*, vol 20, 2011, pp 52–63.

Nozick, Robert, *Philosophical Explanations*, Oxford, Clarendon Press, 1981.

O'Brien J. and Opie J., 'Multiplicity of consciousness and the emergence of self' in *The Self in Neuroscience and Psychiatry*, eds. Tilo Kirscher and Anthony David, CUP, 2003.

Oderberg, David S., 'Comments and Criticisms, Johnston on Human Beings', *The Journal of Philosophy*, 1989, LXXXVI, no. 1, pp. 137-141.

Oppenheimer L., 'Self or selves?: dissociative identity disorder and complexity of the self-system,' in *Theory & Psychology*, vol 12, no.1, pp. 97-128, 2002.

Panksepp Jaak and Northoff Georg, 'The trans-species core SELF: The emergence of active cultural and neuro-ecological agents through self-related processing within subcortical-cortical midline networks,' *Consciousness and Cognition*, 2009, vol. 18, pp. 193–215.

Papineau D., 'Mental Disorder, Illness and Biological Disfunction' in *Philosophy* 37, 1994.

Parfit, D., 'Personal Identity', *Philosophical Review*, 1971, 80, pp.3-27.

Parfit, D., *Reasons and Persons*, Clarendon Press, Oxford, 1984.

Parfit, D., 'Divided Minds and the Nature of Persons' in *Mindwaves*, C. Blakemore and S. Greenfield (eds.), Basil Blackwell, Oxford, 1987.

Parfit, D., 'The Indeterminacy of Identity: A Reply to Brueckner', *Philosophical Studies*, 1993, 70, pp.23-33.

Parfit, 'The unimportance of identity,' in *The Oxford Book of the Self*, edited by Sean Gallagher, OUP, 2011.

*Pavilion: voices of plurality in action*, <http://www.karitas.net/pavilion/>

Perry, John (ed.), *Personal Identity*, University of California Press, California, 1975.

Perry, John, 'The Importance of Being Identical', in *The Identities of Persons*, Amelie Rorty (ed.), University of California Press, California, 1976.

Perry John, 'The Problem of the Essential Indexical,' *Nous* vol. 13, no1, 1979, p. 3-21.

Pinker S, *The Language Instinct*, Penguin Books, London, 1994.

Piper A. and Merskey H., 'The Persistence of Folly: A Critical Examination of Dissociative Identity Disorder, Part I: The Excesses of an Improbable Concept,' *Can J Psychiatry*, 2004, vol. 49, pp592–600.

Piper A. jr., 'Multiple personality disorder', *The British Journal of Psychiatry*, vol 164, 1994, pp 600-612.

Polkinghorne, 'Narrative and Self-concept', *Journal Op Narrative And Life History*, 1(2 & 3), 1991, pp.135-153.

Praetorius, Nini, 'The phenomenological underpinning of the notion of a minimal core self: A psychological perspective,' *Consciousness and Cognition*, 2008, doi:10.1016/j.concog.2008.03.005

Proust Joelle 'Thinking of Oneself as the Same', in *Consciousness and Cognition*, vol 12, pp 495-509, 2003, Elsevier.

Ramachandran V.S. and Blakeslee S, *Phantoms in the Brain: Probing the Mysteries of the Human Mind*, Quill William Morrow, 1998.

Reed, Edward and Jones, Rebecca (eds), *Reasons for Realism: selected essays of J.J.Gibson*, Lawrence Erlbaum Assoc., Hillsdale, N.J, 1982.

Reichenbach, Hans, *Experience and Prediction*, Phoenix Edition, University of Chicago Press, 1961.

Rescher Nicholas, *Process Philosophy: a survey of basic issues*, University of Pittsburgh, Digital Research Library, 2009, <http://digital.library.pitt.edu/cgi-bin/t/text/text-idx>.

Rey, George, 'Survival' *The Identities of Persons*, Amelie Rorty (ed.), University of California Press, California, 1976.

Rey, George, *Contemporary Philosophy of Mind*, Blackwell Publishers, Massachusetts, 1997.

Ricouer, P., *Time and Narrative*, trans. Kathleen Blamey, University of Chicago Press, Chicago, 1984.

Ricouer, P., *Oneself as Another*, trans. Kathleen Blamey, University of Chicago Press, Chicago, 1992.

Rochat, P, 'The self as phenotype,' *Consciousness and Cognition*, Vol. 20, 2011, pp. 109-119.

Rorty, Amelie (ed.), *The Identities of Persons*, University of California Press, California. 1976.

Rorty, Richard, 'Holism, Intrinsicity and Transcendence', in *Dennett and His Critics Demystifying Mind*, Bo Dahlbom (ed.), Blackwell, Oxford, 1993.

Rovane, Carol A., 'The Epistemology of First-Person Reference', *The Journal of Philosophy*, 1987, LXXXIV, no. 2, pp. 147-167.

Rudder Baker L., 'Why Constitution is Not Identity,' *The Journal of Philosophy*, Vol. 94, No. 12, 1997, pp. 599-621.

Rudder Baker L., 'Beyond the Cartesian Self,' from a paper presented at *The Cartesian 'Myth of the Ego' and the Analytic/Continental Divide conference*", Radboud University, Nijmegen, 3-4 September, 2010.

Ruminjo A. and Mekinulov B., 'A case report of Cotard's Syndrome,' in *Psychiatry*, Edgmont, vol. 5(6), 2008, pp 28–29.

Rutkowska, J.C., 'Action, Connectionism and Enaction: a Developmental Perspective', *Connectionism in Context*, A. Clark and R. Lutz (eds.), London, Springer-Verlag, 1992.

Sacks O, *The Man Who Mistook His Wife for a Hat and other clinical tales*, Summit Books, USA, 1985.

Sacks O., *Seeing Voices*, Picador, London, 1991.

Sartre Jean-Paul, *Being and Nothingness* (1943), Citadel Press, 2001.

Schechtman Marya, 'The narrative self', *The Oxford Handbook of the Self*, Sean Gallagher (ed), OUP, 2011.

Schechtman, Marva, 'Personal Identity and the Past' in *Philosophy, Psychiatry, & Psychology*, Vol. 12, No.1, March 2005.

Schechtmann Marya, *The Constitution of Selves*, Ithaca, Cornell University Press, 1996.

Scropo J., Drob S., Weinberger J., and Eagle P., 'Identifying Dissociative Identity Disorder: a self report and projective study,' in *Journal of Abnormal Psychology*, 1998, vol. 107, No 2, pp 272-284.

Seager, William, 'Verificationism, Scepticism, and Consciousness', Symposium: Daniel C. Dennett's Consciousness Explained *Inquiry*, 1992, 36, pp. 113-133.

Searle, John, 'Minds, Brains and Programs,' *Behavioral and Brain Sciences*, vol. 3, 1980, pp. 417-457.

Shalom, A and Robertson, J, 'Hartshorne and the problem of personal identity,' *Process Studies*, Vol. 8, no. 3, 1978

Shoemaker, S., *Self-Knowledge and Self-Identity*, Cornell University Press, 1963.

Shoemaker S., 'Self-reference and self-awareness', *Journal of Philosophy*, vol 65, no.19, pp555-567, 1968.

Shoemaker, S., 'Embodiment and Behavior' in *The Identities of Persons*, Amelie Rorty (ed.), University of California Press, California, 1976.

Shoemaker, S., 'Lovely and Suspect Ideas', *Philosophy and Phenomenological Research*, 1993, *LIII*, No. 4, pp. 905-910.

Shoemaker, S and Swinburne, R, *Personal Identity*, Oxford: Basil Blackwell, 1984, pp. 108-110.

Shoemaker, S., 'Parfit on Identity' in *Reading Parfit*, Jonathan Dancy (ed.), Blackwell Publishers, Oxford, 1997.

Shoemaker, S., 'Sydney Shoemaker's review of Strawson's *Selves*,' [Notre Dame Philosophical Reviews](#), 2013.

Shoter, J., 'The Social Construction of Our Inner Lives', *The Journal of Constructivist Psychology*, 1997.

Simons, J, (ed.) *Contemporary Critical Theorists: from Kant to Lacan*, Edinburgh University Press, 2006.

Sloman Aaron, 'The Self' – a bogus concept, pp 1-7, accessed November 2010. <http://www.cs.bham.ac.uk/research/projects/cogaff/misc/the-self.html>,

Smart, J. J., *Philosophy and Scientific Realism*, Routledge and Kegan Paul, London, 1963.

Solomon, Robert C, *Continental Philosophy since 1750, The Rise and Fall of the Self*, Oxford University Press, Oxford, 1988.

Spanos N. P., 'Multiple identity enactments and multiple personality disorder: A sociocognitive perspective', *Psychological Bulletin*, vol.116, 143-165, 1994.

Spender, Dale, *Man Made Language*, Routledge and Kegan, London, 1980.

Sperry R., Gazzaniga M. and Bogen J., 'Interhemispheric relationship: the neocortical commissures; syndromes of hemisphere disconnection,' *Handbook of Clinical Neurology*, vol. 4, 1969.

Spiegel D., 'Letter to the Executive,' *International Society for the Study of Multiple Personality and Dissociation*, vol 11, 1993.

Stone, Tony and Young, Andrew W, 'Delusions and Brain Injury: The Philosophy and Psychology of Belief', *Mind & Language*, 1997, 12, Nos. 3/4, Sep./Dec., pp. 327-364.

Strawson G., *Selves: an essay in revisionary metaphysics*, OUP, 2009.

Strawson G., 'Against narrativity,' *Ratio (new series)* XVII 4 December, 2004.

Strawson G., 'The phenomenology and ontology of the Self' in *Exploring the self*, D. Zahavi (ed), John Benjamins Publishing Company, Amsterdam, 2000.

Strawson G., 'The Self' and 'The Self and the SESMET' in *Models of the self*, eds. S Gallagher and J Shear, Imprint Academic, UK, 1999.

Strawson G., 'The Self and the SESMET,' *Journal of Consciousness Studies*, 1999, vol 6, no 4, pp. 99-135.

Strawson, G, 'The Self', *Journal of Consciousness Studies*, 1997, vol 4, No. 5/6, pp. 405-428.

Strawson P., *Individuals*, London, Methuen & Co Ltd. 1959.

'Tapes raise fresh doubts on "Sybil" case in U.S.', San Fransisco, *Reuters News*, August 17<sup>th</sup>, 1998.

Teichman, J. and Evans, K., 'The Existence and Identity of Persons' *Philosophy: a beginner's guide*, Blackwell, Oxford, 1992.

Temple A., *This is our truth*, 2007, <http://astraeasweb.net/plural/ourtruth.html>, accessed May 2010.

Thomasma D., 'Moral and metaphysical reflections on Multiple Personality Disorder,' *Theoretical medicine*, 2000, vol 21, pp 235-260.



Trehub A., 'Two arguments for a pre-reflective core self: Commentary on Praetorius,' *Consciousness and Cognition*, vol 18, 2009, pp 339–340.

Uddin L., Davies M., Scott A, Zaidel E., Bookheimer S., Iacoboni M and Dapretto M., 'Neural Basis of Self and Other Representation in Autism: An fMRI Study of Self-Face Recognition,' *PLoS ONE*, 2008, Volume 3, Issue 10, pp 1-9.  
[www.plosone.org](http://www.plosone.org)

Uddin L., Iacoboni M, Lange C and Keenan J., 'The self and social cognition: the role of cortical midline structures and mirror neurons' *TRENDS in Cognitive Sciences*, Vol.11 No.4, pp 153-157.

Unger P., 'The mental problems of the many,' *Oxford Studies in Metaphysics: volume 1*, Clarendon Press, Oxford, 2004.

Unger P., 'There Are No Ordinary Things', *Synthese*, vol. 41, 1979, pp 117–54.

Unger P., 'I Do Not Exist' in *Perception and Identity*, ed. G. F. MacDonald, London: Macmillan, 1979, pp. 235–51.

Unger P., 'Why There Are No People,' *Midwest Studies in Philosophy*, vol. 4, 1979, pp 177–222.

V of the Anachronic Army, *Why We Do Not Have MPD/DID*,  
<http://www.astraeasweb.net/plural/v.html>

Van Inwagen P., 'Meta-ontology,' in *Erkenntnis* vol. 48, pp 233–250, 1998.

Van Inwagen P., 'Materialism and the psychological-continuity account of personal identity' in *Philosophical Perspectives, 11, Mind, Causation, and World*, 1997.

Van Inwagen P., 'Symposium on *Material Beings*,' *Philosophy and Phenomenological Research*, vol 53, 1990, pp 683-719.

Van Inwagen P., *Precis of Material Beings\** in *Philosophy and Phenomenological Research* , Vol. 53, no. 3, September 1993.

Varela, F., Thompson, R., and Rosch, E., *The Embodied Mind*, MIT. Press, Cambridge, Mass, 1992.

Velleman, D J., 'The Self as Narrator', Department of Philosophy, University of Michigan, 1999, [URL:http://country.rs.itd.umich.edu/~veleman/self/narrator.html](http://country.rs.itd.umich.edu/~veleman/self/narrator.html).

Velleman D J, *Self to Self: selected essays*, Cambridge University Press, New York, 2006.

Velleman D J, 'So It Goes,' The Amherst lectures in philosophy: Lecture 1, 2006,

<http://www.amherstlecture.org/>

Vogeley Kai and Gallagher Shaun, 'Self in the Brain,' p.110, in Gallagher S. (ed), *The Oxford Book of the Self*, OUP, 2011.

WebMD, LLC, *Dissociative Identity Disorder*, 2005-2013,' <http://www.webmd.com/mental-health/dissociative-identity-disorder-multiple-personality-disorder>.

Weisberg J., 'Misrepresenting consciousness,' *Philosophical Studies*, Springer, 2010.

Weisberg J., 'Being all that we can be: a critical review of Thomas Metzinger's Being no-one,' *Journal of Consciousness Studies*, vol. 10, no. 11, 2003, pp 89-96.

Wilkes K., 'Know Thyself,' in *Models of the self*, eds. S Gallagher and J Shear, Imprint Academic, U.K., 1999.

Wilkes, Kathleen V., 'The Relationship Between Scientific Psychology and Common-sense Psychology', *Synthese*, 1991, Oct., pp.15-39.

Wilkes, Kathleen V., 'Multiple Personality and Personal Identity', *British Journal for the Philosophy of Science*, 1981, vol 32, pp.331-348.

Williams, Bernard, 'The Self and the Future', *Personal Identity*, John Perry (ed.), University of California Press, California, 1975.

Winnicott Daniel, *The maturational processes and the facilitating environment: Studies in the theory of emotional development*, Hogarth, London, 1960.

Wittgenstein L., *Philosophical Investigations*, revised edition P.M.S. Hacker and Joachim Schulte, Blackwell Publishing, U.K., 2009.

Young, Andrew W, *Face and Mind*, Oxford University Press, Oxford, 1998.

Young Kay and Saver Jeffrey, 'The Neurology of Narrative' in *Substance* vol. 94/95, pp 72-84, 2001.

Zahavi D., 'Self and other: The limits of narrative understanding' in D. D. Hutto (ed) *Narrative and Understanding Persons*, Cambridge, Cambridge University Press, 2007, pp. 179-201.

Zahavi D., 'Being someone,' *PSYCHE*, vol 11, no 5, 2005, pp 1-20.

Zahavi D. (ed), *Exploring the Self: philosophical and psychological perspectives on self experience*, John Benjamins Publishing Company, The Netherlands, 2000.

Zahavi D and Parnas J, 'Phenomenal consciousness and self-awareness: a phenomenological critique of representational theory,' in *Models of the Self*, eds. S Gallagher and J Shear, Imprint Academic, U.K., 1999.

Zimmerman E., (ed) *Oxford Studies in Metaphysics: volume 1*, Clarendon Press, Oxford, 2004.

# APPENDIX

## Letter from Dennett

Yvonne Egege  
Philosophy Dept., Flinders University  
Bedford Park, Adelaide  
S. Australia  
June, 1995

Dear Yvonne

I have finally found the time to read your essay, which is very good, and send you a few comments. I hope they are not too late to be useful.

Overall, you've done an excellent, nearly flawless, job of interpreting and analysing my position, and I think you are right in your general claims and conclusions. I did underestimate (or at least underplay) the importance of (1) the biological self in providing the basis for the social or selfy-self, and (2) the factual nature of the narrative we spin (or that spins us).

Now in part you create a greater distance than really exists – I take most of what you say to be a congenial re-expression with different emphasis, rather than a criticism – but the fact remains that my texts permit and even encourage the interpretation your criticize, so it's fair game.

A few objections

p. 1. 'nonreductive in that they cannot be explained by cognitive function alone.' That is too strong. Cognitive function is (more or less) what can explain the self. Underlying biology, however, is what the self cannot be reduced to.

My claims about where there is no fact of the matter are more restricted than you suggest. It is only rather special borderline cases where I say there is no fact. In a year or so there will be an issue of Phil. Topics devoted to my work. Rosenthal and Block have essays that make the same mistake you do, and I reply to them at length on this score.

You treat 'concrete' and 'real' as too close to synonymous. I take virtual machines and centres of gravity and, say, Australian democracy to be perfectly real, but also (Don't ask: what it is made of, and how much does it weigh, etc.) So when you say 'real rather than abstract' (p. 5) or emphasise the reality (concreteness?) of the biological self (p. 7), I don't think you are dealing a direct blow to anything I maintained. The biological self is not like, say, the nucleus of a cell, or the gall bladder, or even the (distributed, poorly bounded) immune system. It is more like health, or homeostasis. Don't you agree? Then even the biological self is not an organ, or even a single system like the immune system, though the proper functioning of such systems is what constitutes the biological self.

p.7. bottom: You have me claiming that 'The selfy-self does not emerge out of the biological self but is a product of social forces...' I would say: The selfy-self does emerge out of the biological self, largely as a product of social forces.' (An instance of you straining to create more disagreement than there really is.

p. 12. "It supposedly presents a unified identity to the outside world...the blunt fact of the singularity of the organism itself.' I disagree. Consider such familiar queries as why on earth did you do that? The singularity of the organism is often confronted with baffling multiplicity (or complexity) of behavior. That's where the Centre of Narrative Gravity comes into its own.

p. 16. The meme, like the gene, acts automatically, but not in any other sense autonomously. Our behavior (and only ours) is not just 'automatic'. But it is still the function of body-host plus memes plus current perception, I claim. I think you agree. (Elbow Room has more on this side of my view.)

Finally, it occurs to me that the way to clarify and emphasise your distinction between self and identity would be to ask yourself what are the implications for your view are for such philosophical fantasies as 'body transplants.' I and others have often imagined moving the information (or the software) constituting one self from one body (and brain) to another, thereby affecting a body transplant. Wouldn't this be much more problematic on your view, than philosophers have often supposed?

Best Wishes, Daniel C. Dennett