THE UNIVERSITY *of* EDINBURGH

The Concept of Self

Thinking of oneself as a subject of thought

Alisa Mandrigin

Submitted for the degree of PhD by research

The University of Edinburgh

2012

I have read and understood the University of Edinburgh guidelines on plagiarism. This thesis was composed by me and is entirely my own work except where I indicate otherwise by use of quotes and references. No part of it has been submitted for any other degree or professional qualification.

Alisa Mandrigin

Acknowledgements

I would like to thank my supervisors Matthew Nudds and Tillmann Vierkant: I have benefited more from their advice and support than I can do justice to with my words of thanks. I am appreciative of all that I have learnt from them.

Many thanks to the Department of Philosophy at the University of Edinburgh, in particular to the postgraduate philosophy students and especially to Mog Stapleton.
I owe special thanks also to Dan Zahavi for welcoming me to the Center for Subjectivity Research for two months in 2010, and to Evan Thompson for hosting me twice at the University of Toronto, in 2008 and 2010.

My doctoral studies have been supported by the University of Edinburgh in the form of the Shaw Philosophical Fellowship 2007/2008, by the AHRC Doctoral Award 2008/140004, and by the Royal Institute of Philosophy through the Jacobsen Fellowship 2010/2011.

I am indebted to my parents, Mikhail and Susan, and to my sister Darya for their extraordinary personal support, but to none more so than Graeme Yule.

Abstract

We can think about ourselves in a variety of ways, but only some of the thoughts that we entertain about ourselves will be thoughts which we know concern ourselves. I call these first-person thoughts, and the component of such thoughts that picks out the object about which one is thinking—oneself—the self-concept.

In this thesis I am concerned with providing an account of the content of the self-concept. The challenge is to provide an account that meets two conditions on first-person thought. The account must show how we are aware of ourselves when we entertain first-person thoughts, so that we have an account that establishes the cognitive significance of first-person thoughts. But, in addition, this awareness must be as robust as the thinker's ability to entertain first-person thoughts if our account is to respect the guaranteed referential success of the self-concept. I introduce both the subject matter of the thesis, and the constraints on a satisfactory account of that subject matter in the first chapter.

In the second chapter I then set up a further problem: much of our self-knowledge is knowledge of our current mental states and it is often argued that we know about and can ascribe those mental states on the basis of introspection alone. The first constraint on an account of first-person thought described in the preceding paragraph requires that we be aware of ourselves in some way if our thoughts are to have the special cognitive significance of first-person thoughts. Yet, I argue, we neither do nor can introspectively observe a subject of thought and experience when we come to know about our mental states and experiences.

The failure of introspection to supply us with perceptual information about a subject of thought presents us with the further potential problem. According to Fregean semantics sense determines reference: we count on the content of the elements of thought to determine the reference of terms that are used to express those elements. If we do not introspectively observe a subject of thought then we seem to be at a loss to account for the concept and we are at risk of having to accept that neither the self-concept nor the first-person pronoun are referential.

In the remainder of my thesis I consider various responses that we can offer to this problem. First, I examine whether we can avoid the problem with an alternative account of first-person reference according to which reference is fixed by a reflexive rule, and whether we can also base an account of first-person thought on this account of first-person reference. Secondly, I look at the descriptivist view of first-person thought which could potentially provide both an account of first-person thought and first-person reference. These two suggestions must be rejected on the grounds that they fail to accommodate the special cognitive significance of first-person thought. A third approach to first-person thought argues that we employ an objective self-concept when we think about ourselves, a concept that is informed by bodily experience, rather than by introspective observation of a subject. Yet such an account cannot make sense of first-person thoughts in which we question our own embodiment. Lastly I consider whether it is possible to explain the cognitive significance of first-person thought in terms of non-conceptual first-person contents.

There was a young man who said, “Though
It seems that I know that I know,
What I would like to see
Is the I that knows me
When I know that I know that I know.”
- Unknown

Contents

Chapter 1: Thinking about oneself 1

1.1 Introduction 1

1.2 The irreducible cognitive significance of first-person thought 2

1.3 Guaranteed reference 9

1.4 Outline of the thesis 14

Chapter 2: Introspectionism and the elusiveness of the self 18

2.1 Introduction 18

2.2 Introspectionism 19

2.3 Hume: we do not observe a self 25

2.4 Kant et al.: we cannot observe a self 31

2.5 First-person thought and perceptual-demonstrative thoughts 35

2.6 The no-reference view 37

2.7 Conclusion 39

Chapter 3: The self-reference rule and self-consciousness 41

3.1 Introduction 41

3.2 Indexicality and the self-reference rule 42

3.3 The self-reference rule and the content of thought 45

3.4 Referential success 48

3.5 The self-reference rule and semantic adequacy 51

3.6 Self-consciousness and first-person thought 55

3.7 First-person reference and first-person thought 58

3.8 Conclusion 59

Chapter 4: Descriptivism and first-person thought 61

4.1 Introduction 61

4.2 Descriptive sense 62

4.3 A description of the subject of thought 66

4.4 Sense, reference and the modal objection 70

4.5 The problem of the proliferation of self-concepts 73

4.6 The cognitive significance of 'I'-thoughts 75

4.7 Conclusion 81

Chapter 5: An objective self-concept 82

5.1 Introduction 82

5.2 Evans's perceptual demonstrative account 84

5.3 An objective self-concept 88

5.4 The potential failure of our perceptual faculties 92

5.5 Dispositions and circularity 94

5.6 Sophisticated thinkers and Bermúdez on the non-conceptual point of view 97

5.7 Thoughts about disembodiment 100

5.8 Conclusion 104

Chapter 6: Fully self-conscious thought 105

6.1 Introduction 105

6.2 Accounting for fully self-conscious first-person thought 105

6.3 Peacocke's account of fully self-conscious thought 110

6.4 Two objections of circularity 115

6.5 Peacocke's account of non-conceptual first-person content 124

6.6 Peacocke and Evans on first-person thought 127

6.7 The metaphysics of mind and a satisfactory account of first-person thought 131

6.8 Conclusion 139

References 142

Chapter 1: Thinking about oneself

1.1 Introduction

We tend to think we know a lot about ourselves. Some of the things we can know about ourselves include our pasts, the state of our bodies, and the state of our minds. Self-knowledge usually refers to the knowledge we have of our mental states, such as our beliefs, hopes, desires and experiences. One can know what one believes; what one hopes; what one desires; what one experiences.

But self-knowledge is not merely knowledge of the existence of mental states and conscious experiences unattributed to a person. When I attribute a mental state to myself, I make reference to myself. When I become aware of a thought through introspection, I recognize that *I* have been thinking, and not just that there is an episode of thinking (Gertler, 2011, p. 209). We can understand self-knowledge as knowledge of the subject of the mental states and events that we learn about through introspection. Self-knowledge is also knowledge of the thing that thinks, believes, and experiences. We can see the significance of this if we consider how it is that we know things.

What do we know? We may know how to do lots of different things. This is usually referred to as know-how, or ability knowledge. For example, one can know how to ride a bike—one knows how to perform the actions requisite to ride a bike, but it would be difficult for one to specify what those actions are. I might also know of individuals, places and particulars, in the sense that I know my friend Carmen and I know London, but I don't know Barack Obama and Washington, D.C. Call this acquaintance knowledge.

We can distinguish these two forms of knowledge from another kind: propositional knowledge, or knowledge of a proposition. A proposition is a sentence or thought that describes a situation. A situation that obtains is described, therefore, by a true proposition. In describing a situation, a proposition will describe some part of the world as being a particular way, or as instantiating a particular property. In this way, we can think of the propositions that we believe to have a subject-predicate structure. In believing a proposition I believe of something that it instantiates a particular property. For example, I will think that Carmen lives in London, or that the Eiffel Tower is in Paris.

Self-knowledge is a form of propositional knowledge. When one has self-knowledge the proposition that one believes will be a proposition about oneself: the proposition that one instantiates a particular property. For us to have self-knowledge, we must believe

propositions about ourselves. To have a belief about oneself, one must entertain a thought about oneself (Bermúdez, 1998, p. 2). Hence, self-knowledge seems to depend on our being able to think about ourselves.

Of course we do think about ourselves, and do so often in ordinary, non-philosophical circumstances. In thinking about ourselves we predicate various different properties to ourselves, both material properties, and psychological or mental properties. In thinking about ourselves, we seem to be doing much the same as when we think about other individuals, and about inanimate objects. I ascribe properties to myself, and I ascribe properties to people and objects. At least a great number of these properties also seem to be common to both our self-ascriptions and our other-ascriptions. I can ascribe being a certain height to myself and to a grandfather clock, and I can ascribe being angry to myself and to another person.

We can think about ourselves in a number of ways. When I think that I am five feet seven inches tall, my belief will have the content that I am five feet seven inches tall. When I think that I am thinking about the self, my belief will have the content that I am thinking about the self. In having such thoughts I employ concepts, which are the constituents of the content of my thought. A concept picks out a particular object or property: it provides one with a way of thinking about something. If we want to understand how it is that we think about ourselves, we need to know more about the constituent of thought, or concept which is supposed to represent each of us when we employ it in thought. I will call this concept the self-concept. There are a number of features that belong to thoughts that employ the self-concept, and we will have to respect these features if we are to give an account of the self-concept.

1.2 The irreducible cognitive significance of first-person thought

We can think about ourselves by picking ourselves out in a number of ways. That is, self-reference can take a variety of forms (Bermúdez, 1998, p. 2). For example, I might think that the next person to buy Mog a pint is foolhardy, without knowing that it will be me that will be the next person to buy Mog a pint. My thought is self-referring, even though I am not aware of that fact. Alternatively, like Ernst Mach (1897), I might catch sight of my reflection in a mirror and think about the person whose reflection I see without realizing it is my own reflection. Mach sees his reflection in the mirror of a streetcar, but, not realizing it is his own reflection, he remarks to himself on the shabbiness of the person whose reflection he sees. Before I realize that it is my own reflection I might think that *that person* is particularly

scruffy. In entertaining such a thought I pick myself out demonstratively, using the perceptual information available to me.

These are third-personal ways of thinking about oneself. They are ways in which one can think about other persons: I can think of other people by picking them out by means of a description which they uniquely satisfy, and I can make demonstrative reference to the individuals I perceive. These third-personal ways of thinking about individuals are also ways in which others can think about me. Another person can think that the next person to buy Mog a pint is foolhardy, or entertain the thought that *that person* is a shabby pedagogue. But, as long as it is me that is making use of those ways of thinking about the person who happens to be me, they will count as instances of self-reference.

But these ways of thinking about oneself do not express self-consciousness. At least some of the time when we think about ourselves we know that we are thinking about ourselves. The thoughts we have which do express self-consciousness seem to have a special and irreducible cognitive significance (Perry, 1977, 1979; Evans, 1982; O'Brien, 2007). I will call thoughts that are expressive of self-consciousness first-person thoughts or 'I'-thoughts. I will call the first-person element of such thoughts the self-concept.

The point is that we might employ a variety of concepts when we entertain thoughts about ourselves. Each of the concepts I might employ when I think about myself will refer to me, but despite being co-referential, those concepts will differ in their cognitive significance. Only first-person thoughts will have the special cognitive significance required to explain our actions.

Take A.A. Milne's (1926) story about Winnie-the-Pooh walking in the snow and hunting for a Woozle. Milne describes the bear walking around a bush following tracks in the snow. When Piglet asks him what he is doing he says that he is hunting for the animal that is making the tracks. Piglet is interested and suspects that they may be the tracks of a Woozle, so he joins Pooh in walking around the tree. After a bit the bear notices that there are now two sets of tracks in the snow, and surmises that a second Woozle has joined the first one. It is only when Christopher Robin arrives and questions them that the truth of the situation dawns on Winnie-the-Pooh: he has been following his own tracks.

Until he realizes that the tracks in the snow are his own, Winnie-the-Pooh can think about the individual who is making the tracks in the snow. And of course, we know that in doing so there is a sense in which he is thinking about himself, but we also know that he does not know that. Thinking of himself in this third-personal way is not sufficient to motivate Winnie-the-Pooh to act in the manner in which he acts when he knows that he *himself* is making those tracks in the snow. Before he comes to this realization Pooh

continues to hunt for the animal that is making those footprints. When he knows that it is himself that has made the footprints, he stops.

Winnie-the-Pooh's situation is much the same as John Perry's predicament, which he recounts in 'The Problem of the Essential Indexical':

"I once followed a trail of sugar on a supermarket floor, pushing my cart down the aisle on one side of a tall counter and back the aisle on the other, seeking the shopper with the torn sack to tell him he was making a mess. With each trip around the counter, the trail became thicker. But I seemed unable to catch up. Finally it dawned on me. I was the shopper I was trying to catch." (Perry, 1979, p. 3)

What Perry notices is that there is a cognitive difference between thinking about oneself first-personally, as he does when he realizes he is the shopper he has been trying to catch, and thinking about oneself third-personally. When one thinks about oneself third-personally one's thought does not prompt the same kinds of behaviours that a first-person thought does.

Perry entertains a third-personal thought about himself when, prior to his realization that he satisfies the description 'the person who is spilling the sugar',[1] he thinks that the person who is spilling the sugar is making a mess. He thinks about himself, about the person who he happens to be, without being self-conscious. He does not know that he satisfies the description by which he individuates a particular person, himself. His realization that it is he that has been spilling the sugar amounts to a fundamental psychological change in Perry (O'Brien, 2007, p. 5). When Winnie-the-Pooh thinks about the animal that is making the footprints prior to his realization that it is himself that is that animal, he does not think about himself first-personally. The shift from a third-personal way of thinking of himself to a first-personal one marks a significant psychological change for Winnie-the-Pooh. He knows that it is he, himself that is making the footprints he is following. When he knows this, his behaviour reflects the cognitive shift he has undergone.

First-person thoughts have a special cognitive significance because they are self-conscious thoughts, and as such they have a special motivational force. My knowledge that I must submit my thesis by the end of August prompts me to go to the postgraduate office early to work. If instead I thought that the next person to buy Mog a pint must submit a thesis by the end of August, and I do not know that I will be the next person to buy Mog a pint, then I will fail to act in the ways appropriate for someone who has a looming deadline. David Kaplan gives the following example:

"If I see, reflected in a window, the image of a man whose pants appear to be on fire, my behaviour is sensitive to whether I think, 'His pants are on fire' or 'My pants are on fire', though the object of thought may be the same." (Kaplan, 1989a, p. 533)

[1] To pick out Perry uniquely the description should be indexed to a time and place, or should pick out the spilt sugar demonstratively.

My first-person thought that my trousers are on fire will move me to behave in ways appropriate to putting out the flames on my own trousers. But if I only believe of the person that I in fact am that her trousers are on fire, I will act in a different set of ways (Shoemaker, 1986, p. 112). If, as Kaplan's example suggests, I see the reflection of someone in the mirror with his trousers on fire and thereby think that *that* person's trousers are on fire, I will be motivated to look for the person who's reflection I see, and perhaps to assist them in putting the fire out. First-person thought has a special role in the explanation of behaviour. One's beliefs about oneself must be first-personal if they are to play a role in explaining our motivation to act at all, and in explaining our choice of particular kinds of self-interested actions. First-person thought therefore plays a special role in the explanation of behaviour because it is not just thought about a person, but self-conscious thought about oneself (O'Brien, 2007, p. 5).

Typically differences in cognitive significance between co-referential terms are explained by differences in the descriptive or perceptual information associated with the terms (O'Brien, 2007, p. 5). If the thought or proposition expressed by a sentence is identified with the truth-conditions of that sentence, then sentences in which distinct but co-referential terms occur will appear to express the same proposition. For example, utterances of the two sentences:

(1) Superman is Superman

(2) Clark Kent is Superman

within the context of the comic, will have the same truth-conditions. If we identify the propositional content expressed by a sentence with its truth-conditions, then (1) and (2) both express the singular proposition:

{Superman, the property of being Superman}

Yet, to Lois Lane (1) will express something trivial, whereas (2) is non-trivial, and potentially informative. Moreover, assuming that she does not know that the shy reporter she works with is one and the same individual as the superhero, she will assent to (1) and dissent from (2). So we can embed the sentences within propositional attitude ascriptions as follows:

(3) Lois Lane believes that Superman is Superman

(4) Lois Lane does not believe that Clark Kent is Superman

Yet, on the view that propositional content should be identified with the truth-conditions of the sentence, then Lois Lane will appear to hold contradictory beliefs:

(5) Lois Lane believes {Superman, the property of being Superman}

(6) Lois Lane does not believe {Superman, the property of being Superman}

To explain these differences in cognitive significance and the rationality of Lois Lane's differing epistemic attitudes towards (1) and (2) Frege suggests we distinguish between the sense and reference of singular terms. The sense is the 'mode of presentation' of the referent. We can account for the fact that Lois Lane believes (1) but not (2) despite the two sentences having the same truth-conditions by assigning different senses to the names 'Superman' and 'Clark Kent'. Lois Lane thinks of Superman in a different way when she uses the name 'Clark Kent' from the way she thinks of him when she uses the name 'Superman'.

Now sense or mode of presentation of the referent is usually understood in terms of descriptive or perceptual information, as Lucy O'Brien (2007) points out. The sense of the name 'Superman' might be thought to be a definite description, or some set of uniquely satisfied descriptions, which denote the referent of the name, Superman. Alternatively, the mode of presentation of Superman might be some perceptual information about the referent, Superman, of which one is in receipt.

We often do know a great many facts about ourselves. We typically are party to some kind of perceptual information about ourselves—for example, one looks down and sees one's hands. But neither of these kinds of information—descriptive or perceptual—seem able to account for the cognitive significance of first-person thoughts by functioning as a mode of presentation of the referent, oneself. Our first-person thoughts cannot be reduced to a thought involving a definite description, nor a *de re* thought about an individual (again, oneself) that we perceive through ordinary means (Evans, 1982, p. 207).

We might give a direct specification of the thought that Winnie-the-Pooh entertains when he realizes that he has been following his own footprints as so:

(i) Winnie-the-Pooh believes the proposition he would express by saying "I am making the footprints".

So, the direct expression of a first-person thought will be a sentence in which the first-person pronoun occurs (Bermúdez, 1998, pp. 4–5). If we wanted to offer an indirect expression of the bear's thought we might try employing either a definite description or a proper name:[2]

(ii) Winnie-the-Pooh believes that the best friend of Piglet is making the footprints.

[2] Castañeda (1966; 1969) suggests the introduction of a device to allow us to offer reports of first-person thoughts in indirect speech. The device, 'he*', is called a quasi-indictor. It is something like Anscombe's (1994) indirect reflexive pronoun 'he, himself'. The report would then be: Winnie-the-Pooh believes that he* is making the footprints. The quasi-indicator allows us to differentiate between cases in which the thinker does not know that it is himself he is thinking about—when the ordinary pronoun 'he' is employed—and cases in which the thinker would express his thought in direct speech with the first-person pronoun (Bermúdez, 1998, p. 4). But the indirect reflexive pronoun or quasi-indicator can only itself be explained in terms of the first-person pronoun, and so cannot be used as a term that expresses a thought to which our first-person thoughts can be reduced.

(iii) Winnie-the-Pooh believes that Winnie-the-Pooh is making the footprints.

But neither (ii) nor (iii) are equivalent to (i), since they have different truth-conditions from (i). Whilst Winnie-the-Pooh is in fact the best friend of Piglet and 'Winnie-the-Pooh' is the name for Winnie-the-Pooh, the bear might be unaware of these two facts. He might be suffering from amnesia and not know that his name is Winnie-the-Pooh, or he might have forgotten that it is he that is Piglet's best friend. For any other description 'the φ' and any other proper name, the same will be true: the bear might be unaware that it is he that is 'the φ' and that it is he that the name picks out. This is of course just to say that one can fail to know which descriptions one satisfies, and fail to know one's name, and yet be able to entertain first-person thoughts, and perhaps also produce utterances of the first-person pronoun. Descriptive information about oneself is not necessary for one to entertain first-person thoughts (Perry, 1977). But neither is it sufficient, as Perry's example of Rudolf Lingens shows:

"An amnesiac, Rudolf Lingens, is lost in the Stanford library. He reads a number of things in the library, including a biography of himself, and a detailed account of the library in which he is lost. He believes any Fregean thought you think might help him. He still will not know who he is and where he is, and no matter how much knowledge he piles up, until that moment when he is ready to say,

> *This* place is aisle five, floor six, of Main Library, Stanford.
> *I* am Rudolf Lingens." (Perry, 1977, p. 17)

When he is able to entertain descriptive thoughts about the person he happens to be, he is not yet entertaining 'I'-thoughts. Rudolf Lingens might have the belief, based on his reading, that the man in aisle five on floor six of Main Library, Stanford is Rudolf Lingens. If the first-person pronoun is used to express descriptive senses, then the description 'the man in aisle five on floor six of Main Library, Stanford' would be a good candidate for such a descriptive sense. But, even when Lingens has the description available to him, and knows that the individual that satisfies that description bears the name 'Rudolf Lingens', he is not able to utter a sentence containing the first-person pronoun, such as 'I am Rudolf Lingens'. He is capable of entertaining the thought represented by the sentence 'The man in aisle five on floor six of Main Library, Stanford is Rudolf Lingens', but he is not able to entertain the thought that would be expressed by the sentence 'I am Rudolf Lingens'. These two sentences express distinct thoughts, evidenced by the fact that Lingens can take different epistemic attitudes towards them (Perry, 1977). Lingens believes the first thought, but does not believe the second one. Hence, descriptions do not capture the content of first-person thought: they are insufficient to do so.

The insufficiency and non-necessity of the possession of descriptive information about oneself for one to be able to entertain first-person thoughts suggests that it cannot be by thinking of ourselves by any particular description that we entertain first-person thoughts. 'I'-thoughts have a special cognitive value or significance which is lacking in what we might call descriptive thoughts, thoughts for which part of the content comprises a description such as 'the person located at aisle five, floor six, of Main Library, Stanford'. This is apparent both in the difference in epistemic attitudes that we can rationally adopt towards 'I'-thoughts and descriptive thoughts, and also in the immediate implications for action that 'I'-thoughts have and that descriptive thoughts lack.

Perceptual information does not fare much better in providing an explanation of the special cognitive significance of 'I'-thoughts. It is not my knowledge that the person whose reflection I now see in the mirror must submit her thesis soon that prompts me to act in appropriate ways. If the perceptual information I receive about that person was to comprise the content of the self-concept, then we would have no explanation of why 'I'-thoughts motivate us to act in the ways that they do, nor why it is that they are self-conscious thoughts.

Perry (1979, 3) describes the self-concept employed in thought which we express by our uses of the first-person pronoun as an essential indexical. On the grounds that descriptive thoughts lack the motivational force of 'I'-thoughts, he argues that we cannot identify the content of the self-concept with a description. Similarly, since perceptual-demonstrative thoughts lacks the motivational force of 'I'-thoughts, we cannot identify the content of the self-concept with perceptual information. The difference in motivational force between 'I'-thoughts and these other kinds of thoughts suggests that descriptive or perceptual information is insufficient to specify or characterize our self-concepts.

There is a particular kind of thought, which I have been calling a first-person thought or 'I'-thought, which has a special and irreducible role in our cognitive economy. Entertaining a first-person thought prompts one to act in specific ways that third-person ways of thinking about oneself do not. What it suggests is that we cannot reduce first-person thoughts to thoughts that involve reflexive reference which is merely third personal. First-person thought is thought that one knows to be reflexive, and we will therefore have some difficulty in providing an account of the first-person concept.

I will look in more detail at the possibility of reducing the self-concept to perceptual information we have about ourselves, and therefore rendering self-conscious self-reference a form of demonstrative reference in Chapters Two and Five, and I will look further at a descriptivist account of the self-concept in Chapter Four. For now, though, we can recognize

that the cognitive significance of first-person thought—the fact that first-person thought expresses self-consciousness—makes accounting for the self-concept problematic.

1.3 Guaranteed reference

Whilst there are difficulties enough in accounting for the self-concept arising from the fact that the self-concept is employed in thoughts that express self-consciousness, there is also a further challenge: providing an account that preserves this feature of first-person thought, but at the same time safeguards another feature: the exceptional robustness of first-person reference (Anscombe, 1994; Bar-On, 2004; O'Brien, 2007):

"But 'I'—if it makes a reference, if, that is, its mode of meaning is that it is supposed to make a reference—is secure against reference-failure. Just thinking 'I…' guarantees not only the existence but the presence of its referent." (Anscombe, 1994, p. 149)

The guaranteed reference of the first-person is typically associated with expressions of first-person thought in language using the first-person pronoun (Anscombe, 1994). An utterance of 'I' will be immune both from reference failure and from mistaken reference. When one uses 'I' the referent of the term seems to be guaranteed to exist and be appropriately present. Moreover, one seems to be guaranteed to refer to the *right* thing—oneself (Bar-On, 2004, p. 72). And the same also seems to hold for the self-concept (O'Brien, 2007, p. 5). When one employs the self-concept in thought one's use of the concept seems to be guaranteed to refer, and to be guaranteed to refer to oneself. We can contrast this immunity to failure that the first-person seems to enjoy with the kinds of failures that can beset other forms of reference. I might utter a term that fails to pick out any individual, perhaps because I use a description that does not pick out any object or individual ('the present king of France', say). Alternatively, I might use a demonstrative like 'that' when I take my experience to present me with an object, when in fact no object exists at all (Anscombe, 1994, pp. 148–9). These kinds of failure do not seem to befall the first-person pronoun or self-concept.

The apparent guaranteed reference of the first-person pronoun has notably led to the conclusion that the term does not in fact refer. What I have called the guaranteed referential success of the first-person pronoun is interpreted by G.E.M. Anscombe (1994) as security from reference failure. It is not so much that we are guaranteed to get hold of the right object when we use the first-person pronoun, but that we cannot get hold of the wrong object. This reinterpretation paves the way for Anscombe to argue that the first-person pronoun is not a referring expression.

A thinker's being secure against getting hold of the wrong object in thought allows for two possibilities. If when we use the first-person we cannot single out the wrong individual, then either the first-person enjoys guaranteed successful reference (as I have been claiming), or it does not refer at all—there is no 'getting hold' of any individual in thought. The fact that we cannot get hold of the wrong object in thought leads us to think that the first-person is guaranteed to refer to the individual that the thinker intends to single out—in other words, herself (Anscombe, 1994, p. 153). And, of course, always getting hold of the right object is one way in which one can successfully never get hold of the wrong object. But there is another possibility: one can never get hold of the wrong object if one never gets hold of any object at all:

"Getting hold of the wrong object *is* excluded, and that makes us think that getting hold of the right object is guaranteed. But the reason is that there is no getting hold of an object at all." (Anscombe, 1994, p. 153)

It is this that Anscombe thinks is going on with the first-person pronoun and we might extrapolate her arguments to the self-concept. Security from reference failure can be explained in two ways: either, the first-person has guaranteed successful reference, in which case, Anscombe argues, it must refer to a Cartesian Ego; or, the first-person is not a referring expression at all. The conclusion that the first-person refers to a Cartesian Ego is "intolerable" (Anscombe 1994, p. 152), she states, so it must be the case that the first-person does not refer.

Anscombe comes to the conclusion that if the first-person refers at all it must refer to a Cartesian ego by way of a thought experiment. She argues that if we take the first-person pronoun to be a quasi-demonstrative—for which reference is determined by perceptual information—then the guaranteed referential success of the first-person requires our self-perception to be equally resilient. If we can imagine a scenario in which a particular perceptual modality fails, but in which we can continue to think about ourselves and produce tokens of the first-person pronoun, then the modality that has been imagined to fail cannot be responsible for the content of the self-concept and thus the determination of first-person reference. But the problem is that all our ordinary perceptual awareness of ourselves can be imagined to fail:

"[…] now imagine that I get into a state of "sensory deprivation". Sight is cut off, and I am locally anaesthetized everywhere, perhaps floated in a tank of tepid water; I am unable to speak, or touch any part of my body with any other. Now I tell myself, 'I won't let this happen again!' If the object meant by 'I' is this body, this human being, then in these circumstances it won't be present to my sense; and how else can it be "present to" me? But have I lost what I mean by 'I'? Is that not present to me? Am I reduced to, as it were, "referring in absence"? I have not lost my "self-consciousness"; nor can what I mean by 'I'

be an object no longer present to me. This both seems right in itself, and will be required by the "guaranteed reference" that we are considering." (Anscombe, 1994, p. 152)

The only thing that could be unfailingly present in all situations in which we can entertain 'I'-thoughts is something like a Cartesian ego or self, as Anscombe's sensory deprivation tank thought experiment is supposed to show. All that the subject in the sensory deprivation tank is presented with is, it is suggested, herself as a subject of thought. As long as we model first-person reference on demonstrative reference it must be, says Anscombe, that our perception of ourselves is perception of something through introspection alone: a Cartesian ego or subject of thought. Anscombe's argument is a *reductio ad absurdum* of the idea that the first-person pronoun is a referring expression. She argues that if 'I' refers, it must refer to a Cartesian ego. She takes the conclusion as absurd and therefore rejects the idea that 'I' refers at all (Anscombe, 1994, pp. 152–3).

Is Anscombe's argument for the no-reference view compelling? Security from reference failure can be explained in two ways: either, the first-person has guaranteed successful reference, or the first-person is not a referring expression at all. She thinks that taking the first-person pronoun to have guaranteed reference leads to the 'intolerable' conclusion that the first-person refers to a Cartesian ego (1994, p. 152), so it must be the case that the first-person does not refer. When we use 'I' we are not in the business of getting hold of anything in thought.

But Anscombe's argument depends on taking the demonstrative model of reference as the only live option. She seems to assume that if we accept that 'I' is guaranteed to refer successfully rather than fail to refer at all, then it must be that reference is determined by some perceptual information of which the subject is in receipt. But rather than follow Anscombe in this assumption, we might argue that 'I' has guaranteed referential success because reference places no epistemic demands on the subject.

To summarize, the apparent referential security of the first-person pronoun can be explained either by saying that the first-person pronoun is not a referential term at all—one cannot get hold of the wrong object in thought because one gets hold of no object at all (Anscombe, 1994, p. 153)—or alternatively by saying that the term has guaranteed referential success. Taking this latter option presents us with further options. Either we take 'I' to be a quasi-demonstrative and we seem to be forced to postulate a mysterious capacity for unfailing self-perception (O'Brien, 2007, p. 7), and according to Anscombe we must further claim that 'I' refers to a Cartesian ego. Or, alternatively, we offer some alternative account of reference according to which we place no epistemic conditions on reference. In Chapter Three I will argue for this latter position, after examining the argument for the no-

reference view in more detail in Chapter Two. Since Anscombe does not recognize the possibility that reference might be fixed in some other way, she thinks we must choose between demonstrative reference and the no-reference view. But since we have not yet considered what other accounts of first-person reference there might be, we need not yet feel compelled to accept Anscombe's no-reference view.

For the time being we only need note that the guaranteed reference of the first-person does not fit well with the view that perceptual information determines first-person reference. Anscombe's sensory deprivation tank thought experiment can be taken as a challenge to that view: what can a proponent of this view say in response to the claim that our perception of the world can fail whilst our capacity for first-person thought and first-person reference prevails? If first-person thought and reference can continue in independence from our receiving any perceptual information, then it does not seem that perceptual information can comprise the content of the self-concept.

The view that first-person reference is determined by perceptual information can be called the perceptual demonstrative account (O'Brien, 2007). If the perceptual demonstrative account is correct, then it seems that the absence of the relevant perceptual information should render us unable to entertain first-person thoughts, and therefore also unable to produce referential utterances of 'I'. Yet this does not seem to be the case. Uses of the first-person pronoun appear to be immune both to reference failure and to mistaken reference.

One of the ways that our perceptual systems can falter is if they provide us with misinformation, and we can imagine that this might be so even with respect to our perception of ourselves. For example:

"We can conceive of being directly hooked-up, say by transmission of waves in some medium, to the body of another. In such a case we might become aware e.g. of the movements of another's limbs, in much the same sort of way that we become aware of the motion of our own limbs." (Armstrong, 1984, p. 113).

If the perceptual demonstrative view is correct then my utterance of 'I' in this situation should concern the source of the perceptual information I receive, rather than being about me (O'Brien, 2007, p. 39). But this does not seem to be the case. Even were I to receive perceptual information about the world from another person's body, my utterances of 'I' do no seem to fail to refer, and neither do they start to refer to the individual that is the source of the perceptual information. Rather, I still refer to myself.

Our perceptual systems can also fail to present us with any information at all about the world, as Anscombe's thought experiment above illustrates. In her thought experiment Anscombe asks us to imagine a situation in which we have no bodily or sensory information. She thinks that even when in the sensory deprivation tank we will still be able to refer using

the first-person. As such, the types of information that are absent when in the tank cannot underlie our ability for first-person reference. It cannot be the case that we come to identify the object referred to by 'I' by being presented with our bodies. There is no presentation of the body whilst in the sensory deprivation tank. For the proponent of the perceptual demonstrative account such a situation should leave the subject unable to produce referential utterances of 'I'. But, again, the conclusion that the proponent of the perceptual demonstrative account is forced to accept cannot be correct. The absence of proprioceptive information about one's body, say, does not prevent one from producing referential utterances of the first-person pronoun, and nor does it prevent one from employing the self-concept in thought to think about oneself.

We saw above that neither descriptive nor perceptual information seem to do a satisfactory job at explaining the cognitive significance of a thought in which the self-concept is employed, and it therefore seems that neither kind of information can be used to characterize the content of the self-concept. The guaranteed referential success of the self-concept might seem to corroborate that claim, in particular it might seem to support the rejection of perceptual information as a good candidate to characterize the content of the self-concept. For any ordinary perceptual information that I might be party to will be information I receive from channels which might potentially fail either to present me with accurate information, or to present me with any information at all.

But it is not the case that the two features of first-person thought—its guaranteed referential success and the fact that it expresses self-consciousness on the part of the thinker—act together to discount perceptual information, or at least quasi-perceptual information, from being used to characterize the content of the self-concept altogether. The fact that first-person thought is self-conscious thought suggests that we are aware of ourselves in some way when we entertain such thoughts: we must bring some kind of information or awareness to bear when we entertain first-person thoughts. Ordinary types of perceptual information fail to capture the kind of content that first-person thoughts must have if they are to have the motivational force that self-conscious thoughts have. Added to this, our ordinary sensory systems can fail in the ways described above, and therefore an account of the self-concept that characterizes the content of the first-person element of 'I'-thoughts in terms of ordinary types of perceptual information must abandon the idea that the first-person is guaranteed to refer successfully. But, even if we want to respect the guaranteed referential success of the first-person, we still need to give some account of how we are aware of ourselves when we entertain first-person thoughts if we are to account for their cognitive significance.

The guaranteed referential success of the self-concept places an additional constraint on the kind of information that can form the content of the self-concept and the kind of systems which can relay that information: it is going to have to be information that is always available to the thinker, and the system is going to have to be such that it never fails in a situation in which the thinker is able to entertain an 'I'-thought. Anscombe's sensory deprivation tank thought experiment pushes us to accept that many of the ways in which we receive perceptual information about ourselves can fail whilst leaving intact our capacity for first-person thought.

Self-consciousness places an epistemic demand on the thinker, and the guaranteed referential success of uses of the self-concept places constraints on the ways in which that epistemic demand can be met. For now, what we need to note is the difficulty in reconciling (a) the apparent guaranteed referential success of the self-concept with the fact that it (b) expresses self-consciousness and hence has a particular cognitive significance and motivational role in our cognitive economy. To be self-conscious one needs to be aware or thinking of oneself in some special and particular way. We have already seen that specifying what way that is is going to be a problem since the obvious contenders do not capture the significance of genuinely self-conscious reflexive thought. But added to this problem is the additional feature of first-person thought: that the self-concept is guaranteed to refer. My thesis is concerned with providing an account of the content of the self-concept that respects these two features, (a) and (b) above, of first-person thought.

1.4 Outline of the thesis

In my thesis I am concerned with providing an account of the content of the self-concept that respects the features of first-person thought described above. The challenge will be to provide an account that meets the constraints that I have discussed in Sections 1.2 and 1.3. The account must show how we are aware of ourselves when we entertain first-person thoughts, so that we have an account that establishes the cognitive significance of 'I'-thoughts. But this awareness must be as robust as the thinker's ability to entertain 'I'-thoughts if our account is to respect the guaranteed referential success of the self-concept.

Contemporary philosophy of mind, across a variety of approaches, has been increasingly enamoured with the categorization of primitive or basic forms of self-consciousness: the various ways we have of becoming aware of ourselves prior to any reflective thought about ourselves. In my thesis I want to bring into focus once again a more robust form of self-consciousness: reflective self-consciousness. We are reflectively self-conscious when we

ascribe properties to ourselves in truth-evaluable judgements. Such thoughts, which we might label first-person thoughts or 'I'-thoughts, are thoughts in which we think about ourselves as such: we entertain thoughts which we know concern ourselves. It is because we have this knowledge that the thoughts can be counted as a form of self-consciousness. I will have recourse to think about some of the ways in which we can be said to be pre-reflectively or non-conceptually self-conscious in this thesis, but only in so far as they elucidate our capacity to entertain first-person thoughts.

In Chapter Two I will look at what might seem the most plausible place to start: introspection. Much of our self-knowledge is knowledge of our current mental states, and it is often argued that we know about those mental states through introspection alone. Indeed, it seems that we can ascribe those mental states to ourselves on the basis of introspection alone. We have seen that when we entertain first-person thoughts we must be aware of ourselves in some way or those thoughts will not be self-conscious thoughts, but we have also seen that ordinary perception cannot provide the content of the self-concept. In this chapter I will consider whether we *introspectively* observe a self or subject of thought when we become aware of our mental states through introspection. Exteroceptive perception of our bodies cannot help us to give an account of the self-concept, but introspective perception of the subject of thought might be reliably present when we entertain first-person thoughts. The problem with such an account is perhaps obvious and is certainly established in the literature. David Hume famously argues that we do not seem to observe any self when we introspect. Moreover, it seems that, by necessity, we could not introspectively observe ourselves in the way that the view requires: subjects of thought and experience are not the kinds of things that can be the objects of their own awareness.

The failure of introspection to supply us with perceptual information about a subject of thought presents us with a further potential problem. According to a Fregean semantic theory sense determines reference: we count on the content of the elements of thought to determine the reference of terms that are used to express those elements. That is, the self-concept will be expressed by utterances of the first-person pronoun, and the referentiality of both concept and linguistic term will depend on the concept being content-ful. If we do not introspectively observe a subject of thought, and in addition we have already seen that other kinds of perceptual information and ordinary kinds of definite description cannot be used to characterize the content of the self-concept, then we seem to be at a loss to account for the concept, and we are at risk of having to accept that neither the self-concept nor the first-person pronoun are referential.

Chapter Three offers a response to the conclusion of Chapter Two, that the first-person is not a referring term or concept. I argue that the fact that the first-person is an indexical indicates that we need to take a different approach from the traditional Fregean one in order to account for the reference of the term. According to the self-reference rule account of the first-person, first-person reference is fixed by the rule that a token of the first-person refers to the individual who produced it. The rule fixes the reference of both the first-person pronoun and the self-concept, so the referent of a token of the first-person pronoun will be the speaker who uttered it, and a token of the self-concept in thought will refer to the thinker of the thought in which it occurs. The self-reference rule account of first-person reference allows us to argue that both the self-concept and the first-person pronoun refer, and, in particular, helps to explain why both seem to enjoy guaranteed referential success. But the rule that fixes reference cannot be all that there is to say about the self-concept. We will not be able to use the first-person pronoun or the self-concept in accordance with the rule unless we are able to understand that rule, and this will require that we are aware of ourselves in some way. For our account of the self-concept to respect our first feature of first-person thought—that first-person thought has a special cognitive significance—then it must be that we are aware of ourselves in some way when we entertain thoughts in which the self-concept occurs. So, whilst the self-reference rule account provides us with a response to the no-reference view, and allows us to say that both the self-concept and first-person pronoun are referential, we still need to provide an account of the content of the self-concept.

In Sections 1.2 and 1.3 of this chapter I have dismissed descriptive and perceptual information on the grounds that they fail to provide the kind of awareness of ourselves that is needed to explain the cognitive significance of first-person thoughts. In Chapters Four and Five I will look in more detail at the problems that arise for accounts that draw on such information in specifying the content of the self-concept. In Chapter Four I will look at what might seem to be a more appropriate description than the ordinary definite descriptions that might be thought to populate our thoughts about other individuals. That is, rather than an objective way of thinking of ourselves that might be employed by anyone, I will examine whether the self-concept might be characterized by descriptive information about oneself as a subject of thought. According to Robert J. Howell (2006) the content of the self-concept should be thought of as the description 'the thinker of *this* thought' where the demonstrative 'this' refers to the mental state of which one is currently introspectively aware. But whilst Howell appears to have provided a description that is uniquely satisfied by the appropriate individual, this way of thinking of oneself does not capture the self-consciousness that is such an important feature of first-person thought. It will always be possible for someone who

thinks of the thinker of an introspected thought to wonder if it is indeed she that is identical with the individual who satisfies the description.

Chapter Five provides an analysis of a second way in which perceptual information might be used to account for the content of the self-concept. In Chapter Two I will reject introspective observation of the subject of thought as a viable way of characterizing the content of a concept of oneself as a subject. In Chapter Five I will consider whether information that allows one to think of oneself *objectively* might be a better option. Gareth Evans (1982) argues that it is by having a capacity to locate oneself in space that one is able to entertain first-person thoughts. Evans's own view is prohibitively sophisticated, but a more moderate version of the same kind of view can be found in José Luis Bermúdez's *The Paradox of Self-Consciousness* (1998). But, whilst Bermúdez's is more widely applicable than Evans's, both accounts cannot make sense of questions about one's embodiment. It seems coherent to wonder whether one is in fact disembodied, for example by entertaining the first-person thought expressed by the sentence 'Am I disembodied?'. An account that understands the self-concept as an objective concept cannot make sense of such a thought.

Finally, in Chapter Six I will look at how we might amend the self-reference rule account of first-person reference. In Chapter Three I claimed that it could provide an account of first-person reference but that it did not have the resources to explain the self-consciousness of first-person thought. In Chapter Six I will look at Christopher Peacocke's (2010) view of how we might explain the fact that a subject who entertains a first-person thought knows that she is thinking about herself. According to Peacocke the resources to explain the self-consciousness of first-person thought are present in all kinds of active thought. As long as a subject has a grasp of the self-reference rule itself, then we can explain the self-consciousness of first-person thought. According to Peacocke a subject's relation to her own active thoughts is such that she is not just aware of them, but she is aware of them as her own. This relation is spelled out in terms of action-awareness with the content *I am judging that P*. What is missing, then, from the self-reference rule account is the recognition that a subject has this awareness whenever she entertains an active thought. When a subject both knows that she is the producer of a first-person thought, and that a token of the self-concept refers to the producer of that token, then a subject will know that it is to herself that she has referred when she thinks that first-person thought.

Chapter 2: Introspectionism and the elusiveness of the self

2.1 Introduction

The focus of my thesis is the content of first-person thought. In particular, I am interested in the content of the first-person element of first-person thought: the self-concept. In this chapter I will look at a *prima facie* plausible view of the content of the self-concept. Many of our first-person thoughts are identification-free, just as perceptual-demonstrative thoughts are identification-free. Moreover, we can ascribe psychological properties to ourselves on the basis of introspection alone, so the suggestion is that the self-concept is a kind of demonstrative term that we use to refer to the self that we introspectively observe.

The idea that the self-concept is introspection-based arises from consideration of Anscombe's sensory tank deprivation thought experiment. Given that one can think about oneself without any experience of one's body or its circumstances, it seems that the awareness of oneself that comprises the content of the self-concept will come from within thought and experience itself. When we think about ourselves in the first-person we employ an introspection-based self-concept. We introspectively observe a subject of thought and it is this observation that provides the mode of presentation of ourselves that we employ in first-person thought.

Such a self-concept would seem to meet the conditions on first-person thought that I set out in Chapter One. The subject of thought, if it can be introspectively observed, would always be present when one introspects one's mental states: it would be observable as the thing doing the introspecting, so an introspection-based self-concept could meet the condition of guaranteed successful reference. Moreover, introspective observation of the subject of thought would provide one with awareness of oneself, so it seems that it might meet the condition of being self-conscious thought also.

Yet introspectionism cannot be right. When we begin to inspect the objects of introspective awareness we soon notice that the self seems to be elusive. The idea that the self eludes introspective awareness can be developed in a number of ways. We can claim that:

- as a matter of fact we do not observe and neither are we acquainted with a self: when we introspect we cannot detect a self; and,
- by necessity we cannot observe the self: any attempt to observe or become acquainted

with the self through introspection will fail to access the subject of thought and experience.

Furthermore, the difference between past-tense memory-based first-person thoughts and past-tense perceptual-demonstrative thoughts, as well as the failure of a demonstrative thought to have the cognitive significance of first-person thoughts give us reason to reject the classification of first-person thought as a kind of demonstrative thought.

If we do not and cannot introspectively observe a self, and, as we saw in the previous chapter, ordinary kinds of perception and definite descriptions cannot comprise the content of first-person thought, then the first-personal way of thinking about oneself appears to be mysterious. The self-concept seems to turn out to be non-referential. And given a Fregean theory of semantics this means that the first-person pronoun will not be a referential term also. These conclusions are unacceptable, and if introspectionism is the only option available to us in giving an account of the self-concept, then we have a problem.

2.2 Introspectionism

In *The Blue Book* (1958) Wittgenstein teases apart what he describes as two uses of the first-person pronoun, laying the foundations for Sydney Shoemaker's distinction between statements that are immune to error through misidentification, and statements that are subject to that error, and thereby setting up an important distinction between two different kinds of thoughts that we can entertain about ourselves (Bermúdez, 1998, p. 5):

"There are two different cases in the use of the word "I" (or "my") which I might call "the use as object" and "the use as subject". Examples of the first kind of use are these: "My arm is broken", "I have grown six inches", "I have a bump on my forehead", "The wind blows my hair about". Examples of the second kind are: "I see so-and-so", "I hear so-and-so", "I try to lift my arm", 'I think it will rain", "I have a toothache". One can point to the difference between these two categories by saying: The cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of an error, or as I should rather put it: The possibility of an error has been provided for." (Wittgenstein, 1958, pp. 66-7)

In examples of the use of the first-person pronoun as object, as Wittgenstein puts it, there is an identification of a person to whom the property is being ascribed. First-person statements of this sort will depend on one having observational knowledge of oneself. I am able to speak about myself in these 'object uses' because I have observational knowledge of a body—my body. Wittgenstein distinguishes these statements from those that do not seem

to involve the identification of a person: subject uses of the first-person. In uses 'as subject' we seem to be referring to a subject of experience: we ascribe experiences and thoughts to ourselves, and thereby seem to be treating ourselves as the subjects of those states. But what makes Wittgenstein remark on them is that he thinks that they do not involve the identification of a particular person:

"On the other hand, there is no question of recognizing a person when I say I have toothache. To ask "are you sure that its *you* who have pains?" would be nonsensical." (Wittgenstein, 1958, p. 67)

According to Shoemaker, Wittgenstein's remarks indicate a retreat from the idea that self-knowledge is infallible to the more moderate claim that some first-person statements are immune to error due to the misrecognition of a person (Shoemaker, 1968, p. 556). A subject is immune to error through misidentification relative to the first-person when it cannot happen that she is mistaken in saying 'I am *x*', because, although she does know of someone that is *x*, she is mistaken in thinking that person to be herself. To make an error through misidentification the subject will mistakenly judge 'I am *x*', whilst having knowledge that some particular thing, *a*, is *x*. The mistake arises from taking '*a*' to refer to the thing that she knows to be *x* (Shoemaker, 1968, p. 557).

What Wittgenstein and Shoemaker highlight is a special kind of epistemic security. It is often claimed that self-knowledge is especially secure, epistemically. The stronger claim of epistemic security is that we are infallible about our own mental states, but this is implausible. As Shoemaker suggests, we can think of immunity to error through misidentification as a weaker kind of epistemic security: when making certain first-person statements, including introspection-based first-person statements, one cannot be mistaken about who it is that instantiates the property being ascribed. Immunity to error through misidentification therefore might be a way of explaining what is special about self-knowledge without resorting to the claim that we cannot be mistaken in the attribution of mental states to ourselves.

Shoemaker presents immunity to error through misidentification as a feature of certain kinds of statements. In the following I will apply his claims to propositional thought. This involves treating propositional thought as compositional, with a subject-predicate structure akin to the grammatical structure of subject-predicate sentences.[3] My thoughts about objects have two components: an individuation component, and a predication component. The

[3] I do not take the claim that thought is compositional in this way to commit me to the claim that the constituents of thought are physically realized in the brain as the Language of Thought Hypothesis maintains (Evans, 1982, pp. 100-101).

individuation component will individuate or pick out the object that the thought is about, whilst the predication component will pick out the property being ascribed to that object.

Many thoughts can be false because there has been an error on either side: the predication may be mistaken, or the identification of the object may be mistaken. Hence my judgement can be false because of an error of mispredication, or an error of misidentification (Kriegel, 2007). For example, if I see a woman wearing a scarf and judge that my sister is wearing a purple scarf, my judgement could be false in two distinct ways. It may be that I am wearing tinted glasses or that the lighting conditions are not optimal, and so the scarf I judge to be purple is in fact blue. If I have made this mistake my error will be an error of mispredication. Alternatively, it may be that I have misidentified the woman I see as my sister. In that case I would have made an error of misidentification. At the opposite extreme, if I cannot be in error at all in making a judgement I would be immune to both errors of misidentification and mispredication.

The distinction between thoughts that are immune to error through misidentification and that involve a use of the self-concept 'as subject' on the one hand, and thoughts that are vulnerable to error through misidentification and that involve a use of the self-concept 'as object' on the other, maps onto a distinction between two kinds of first-person thought. This is the distinction between propositional thoughts that can be evaluated into two more basic thoughts and thoughts that cannot (Bermúdez, 1998, p. 5). A thought that is *vulnerable* to error through misidentification will (typically—see footnote 10 below) be one that can be broken down into two component thoughts, whilst one that is *immune* to error through misidentification will be such that it cannot be analyzed in this way.

Take the thought I have about myself—that I have a black eye—which I entertain on the basis of seeing the reflection in the mirror of a person with a black eye. And suppose that I am not mistaken in ascribing the property of having a black eye to someone. My thought can be broken down into two more basic thoughts: the thought that *that* woman has a black eye, and the thought that *that* woman is me. The composite thought that I have a black eye will be susceptible to error through misidentification: it might seem unlikely, but I can be mistaken in thinking that the woman I see reflected in the mirror is me. Call thoughts of this kind identification-dependent (Evans, 1982, p. 180). Thoughts that are identification-dependent will be susceptible to error through misidentification. A thought will be *subject* to error through misidentification if one has made an erroneous identity judgement. A thought will be immune to error through misidentification if the thought is not composed or cannot be analyzed in terms of more basic thoughts, one of which involves an identity judgement. Call

thoughts of this kind identification-free thoughts (Evans, 1982, p. 180).

As Shoemaker (Shoemaker, 1968; Shoemaker & Swinburne, 1984) and Bermúdez (1998) point out, one's entertaining a first-person thought that is identification-dependent will depend on one's already having the capacity to entertain first-person thoughts that are identification-free. Any identification-dependent first-person thought will contain an individuation component that is an identity judgement of the form 'I am *a*'. This identity judgement, itself a first-person thought, will employ the self-concept. If the identity judgement that is the individuation component of the first thought is itself identification-dependent, it will also contain an identity judgement as individuation component. And then we must ask the same question of this second identity judgement: is the second identity judgement identification-free or identification-dependent? On pain of infinite regress we must, at some point, come to a first-person thought that is identification-free (Bermúdez, 1998, p. 7).

One of the kinds of first-person thought that is identification-free, and perhaps the paradigm case of a thought that is identification-free, is introspection-based first-person thought. Shoemaker's claim that some thoughts that involve the attribution of "predicates which I apply to others, and which others apply to me, on the basis of observations of behaviour, but which I do not ascribe to myself on this basis" (Shoemaker, 1968, p. 562) are immune to error through misidentification is the claim that thoughts that involve the predication of mental states to oneself on the basis of introspection can only be false because of an error of mispredication, and not because of a misidentification. They are therefore thoughts that are identification-free and that cannot be decomposed or analyzed into two distinct more basic thoughts.

However introspection-based first-person thoughts are not the only kind of first-person thoughts that are identification-free (Evans, 1982), and neither are first-person thoughts the only kind of thoughts that can be identification-free. Perceptual-demonstrative thoughts seem to be thoughts that are identification-free.[4]

Observation of an object allows us to think about an object in (at least) two ways: either we individuate the object demonstratively, or we think about the object by identifying the object individuated demonstratively with an object we know about in some other way. This latter way will be a form of identification-dependent thought. For example, if I see a blue book and think that *The Varieties of Reference* is blue, my thought can be analyzed into the

[4] I will look at the identification-freedom of certain of our self-ascriptive thoughts of bodily properties in greater detail in Chapter Five.

more basic thoughts that *that* book is blue, and that *that* book is *The Varieties of Reference*.

But suppose that I entertain the thought that *that* book is blue when I am in the circumstances described above. We might think that my perceptual experience justifies me in believing that the book is blue. But, we might also think that I am able to entertain the thought at all because of my perceptual experience of the book. It is by perceiving the book that I am able to think about the book. I think about the book demonstratively as *that* thing, and the content of my perceptual experience supplies the means for the individuation of the book in thought: my perceptual experience of the book is a particular mode of presentation of the book. Thoughts of this kind are often called *de re* thoughts: they are thoughts that are object-dependent since if the object—or individual—did not exist, then the thought would not exist either. And to entertain a *de re* thought the subject must stand in some kind of epistemic relation to the object about which she thinks.[5] Now, a thought is identification-free if it cannot be analyzed in terms of more basic thoughts, one of which involves an identity judgement. My demonstrative judgement that *that* book is blue appears to be just such a thought. I do not identify a particular object independently of my ascription of the property of being blue. Rather I think of *that* thing, that it instantiates the property of being blue. So perceptual-demonstrative judgements are identification-free. It might seem suitable, then, to model first-person thought on perceptual-demonstrative thought. First-person thoughts that are identification-free are perceptual-demonstrative thoughts, and first-person thoughts that are identification-dependent will depend, at some point down the line, on a perceptual-demonstrative individuation of oneself. But, what kind of perception might enable us to entertain identification-free first-person thoughts?

We can think about ourselves in the same kind of way that we think about other individuals and objects. I ascribe the property of being broken to my arm when I visually perceive an arm that is broken in the sort of place that my arm usually occupies. Say that I entertain the thought that *that arm* is broken. Some of our thoughts about ourselves, then, seem to be based on ordinary kinds of perceptual experience, and it is the content of the experience that provides the mode of presentation of the object we are thinking about. The content of my visual experience of my arm provides the mode of presentation of the arm that comprises part of the content of my thought about it. As with the perceptual-demonstrative thought that *that* book is blue, the thought will be identification-free. Could it be that the content of first-person thought is characterized by ordinary perceptual experience of our

[5] There is a great deal of discussion about how to understand *de re* belief. I will not offer a defence of my understanding of the phenomenon, but merely assume the view I have described here.

bodies from the outside?

It does not seem likely. To render a thought about one's arm being broken a first-person thought one needs to identify the arm that is seen as one's own arm. The ordinary perceptual-demonstrative thought that *that* arm is broken is indeed identification-free, but the first-person thought that we would express with the sentence 'My arm is broken' is identification-dependent just as my judgement that I have a black eye is identification-dependent. To entertain a first-person thought about one's arm on the basis of ordinary perception of one's body from the outside one will have to make an identity judgement, and the resulting first-person thought will contain that identity judgement. The identification-free thought that *that* arm is broken will not be a self-conscious thought.

Moreover, many of our thoughts about ourselves are not made on the basis of ordinary kinds of perceptual experience of our bodies. Often we think about ourselves without having any exteroceptive perceptual experience of our bodies. The fact that we can do so might lead us to think that it is experience of our bodies 'from the inside'—proprioceptive experience of the body—that comprises the content of first-person thought in such situations.[6]

But now recall Anscombe's thought experiment, which I used in the previous chapter to raise some initial questions about the referentiality of the first-person pronoun and self-concept:

> "[…] now imagine that I get into a state of "sensory deprivation". Sight is cut off, and I am locally anaesthetized everywhere, perhaps floated in a tank of tepid water; I am unable to speak, or touch any part of my body with any other. Now I tell myself, 'I won't let this happen again!' If the object meant by 'I' is this body, this human being, then in these circumstances it won't be present to my sense; and how else can it be "present to" me? But have I lost what I mean by 'I'? Is that not present to me? Am I reduced to, as it were, "referring in absence"? I have not lost my "self-consciousness"; nor can what I mean by 'I' be an object no longer present to me. This both seems right in itself, and will be required by the "guaranteed reference" that we are considering." (Anscombe, 1994, p. 152)

When she imagines herself to be in a sensory deprivation tank, without any sensory experience of the world or her own body, either from the outside or the inside, Anscombe still believes she will be able to entertain first-person thoughts. And this seems correct: there is a way we have of thinking about ourselves that is independent on any kind of awareness of our bodies or our surroundings and circumstances at all. When in the sensory deprivation tank, Anscombe is able to entertain first-person thoughts, but she expressly lacks information

[6] Self-ascriptive thoughts that we entertain on the basis of proprioceptive experience of our bodies seem to be identification-free (Evans, 1982). In Chapter Five I will examine Evans's (1982) claims about the role of information about the body in the individuation of oneself in first-person thought. According to Evans it is necessary for one to employ such a perceptual mode of presentation of oneself if one is to entertain a thought about oneself at all.

about her body, and she also lacks information about the world that would allow her to think about herself relative to the world. On the basis of introspection alone one is able to think about oneself. So, with the perceptual-demonstrative model in mind, we might consider whether it is through introspective observation of a self that we come to think first-person thoughts. When we introspect our mental states we also observe the subject of those states, and it is to that object that we refer when we employ the introspection-based self-concept. I will call this view introspectionism. According to introspectionism the content of the self-concept comprises an introspective-observational mode of presentation of the thinking subject.

Introspectionism appears to meet at least one of the conditions on first-person thought that I discussed in the previous chapter. First-person thought has guaranteed referential success, which suggests that the object which one is thinking about—oneself—must be present so that one can be aware of oneself in the appropriate way whenever one it able to entertain a first-person thought. Since the suggestion is that we think about ourselves by entertaining an observational concept, those first-person thoughts will certainly be thoughts in which one is aware of oneself.[7] And since an introspectively observed subject of thought would be observed when one introspects one's mental states, any circumstance in which one can entertain a first-person thought will be a circumstance in which the subject is present.

However, the problems with introspectionism should be clear: when we introspect, the self seems to elude us. According to David Hume, we do not introspectively observe or become acquainted with a self. According to a Kantian view of the mind, the subject of thought and experience is not the kind of thing that can be observed. Moreover, Shoemaker points out that we have good reason to think that it is a mistake to model first-person thought on perceptual-demonstrative thought. One of the reasons he gives is that the introspectionist's perceptual-demonstrative thought would fail to have the cognitive significance of first-person thought: it would not be thought that is genuinely self-conscious.

2.3 Hume: we do not observe a self

Probably the most famous challenge to the idea that we have introspective awareness of the self can be found in David Hume's *Treatise of Human Nature*, in which he denies the existence of a perceptible, substantive self.

[7] We will see in Section 2.5 that the introspectionist in fact fails to meet the requirement on first-person thought that it be self-conscious thought, since introspection of a subject is not an awareness of oneself as oneself.

In the section titled 'Of Personal Identity' Hume addresses himself to the task of rejecting the suggestion that introspection reveals an observable self.[8] Hume, being an empiricist, takes knowledge of any thing to be based solely on perception of that thing. Hence, knowledge of one's self, even if the self is a thinking thing, could be based only on perception of that self (Atkins, 2005a, p. 34). Yet, when he 'looks within'—when he introspects—he finds nothing but thoughts and experiences.

There are two parts to Hume's argument. The first is the denial that he has any clear *idea* of self.

"[…] nor have we any idea of *self*, after the manner it is here explain'd. For from what impression cou'd this idea be derived? […] It must be some one impression, that gives rise to every real idea. But self or person is not any one impression, but that to which our several impressions and ideas are suppos'd to have a reference. If any impression gives rise to the idea of self, that impression must continue invariably the same, thro' the whole course of our lives; since self is suppos'd to exist after that manner. But there is no impression constant and invariable. Pain and pleasure, grief and joy, passions and sensations succeed each other, and never all exist at the same time. It cannot, therefore, be from any of these impressions, or from any other, that the idea of self is deriv'd; and consequently there is no such idea." (Hume, 2000, T 1.4.6.2, p. 164)

For the empiricist Hume, all knowledge stems from sense impressions. We can only think about things—have ideas of things—that have been encountered in experience (Noonan, 1999, pp. 6-7). Thus, if we are to know the self, then we must perceive it first. Since Hume is concerned with personal identity in this section, he assumes that the idea of the self that we claim to have is an idea of an enduring thing. It is an idea of something that remains the same through time, and it is this idea of self that Hume objects to. To have a sense impression or experience of the unchanging enduring self, that impression would itself have to remain constant and invariable. But our impressions are not like this at all. Our mental states "succeed each other with an inconceivable rapidity, and are in perceptual flux and movement" (T 1.4.6.4, p. 165). As such, Hume professes that he has no clear idea of the self. The self that Hume fails to observe is therefore a substantial, persisting self: a self that could be used to explain how it is that persons endure through time when their experiences and thoughts are fleeting and varied.

The second part of Hume's argument concerns the absence of a perceptible self amongst the objects of introspection.

"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.

[8] In the following discussion I am not concerned with historical scholarship and exegetical debate. I take it that the position I ascribe to Hume is one of the standard interpretations of Hume's argument in 'Of Personal Identity'.

I never can catch *myself* at any time without a perception, and never can observe any thing but the perception." (T.1.4.6.3, p. 165)

We can introspect only thoughts and experiences, and no self to which those states could belong, or of which they could be properties. Hume claims that when we come to know our mental states and conscious experiences we are not presented with an object to which those states and experiences could be ascribed in the way that, when I am aware via visual perception of an object instantiating a particular property, I am able to ascribe the property to the object. If we take the mental states which we come to know about through introspection to be properties of a self, we do not seem to be aware of those mental states as being instantiated in a self. When we introspect we only ever uncover experiences and thoughts. This is called the elusiveness thesis.

Elusiveness thesis (1): *When one introspects and thereby finds out about one's mental states, one does not observe any self to whom those states belong.*

Should we be convinced by Hume's arguments? A great many philosophers have accepted Hume's elusiveness thesis, but it is not without its detractors also. Shoemaker (1986) raises one concern with Hume's project, pointing out that it is not properly empirically based:

"But Hume's denial that he is aware of a self can hardly have the same basis as my present well-founded denial that I see a teakettle. The latter denial is well founded only on the assumption that I have some idea of what it would be like to see a teakettle. Hume, on the other hand, is quite emphatic on the point that he has no idea of "self" (*qua* subject of experiences) and so, presumably, no idea of what it would be like to introspect one. […] First appearances notwithstanding, the basis of the Humean denial can hardly be empirical." (Shoemaker, 1986, pp. 101-102)

Hume claims that he has no idea of 'self', and so it would seem he has no idea of what it would be like to introspect a self. Since he doesn't know what he is searching for, he does not seem to be in a position to deny that he has introspectively perceived a self.

Shoemaker's point concerns the seeming incompatibility of the two parts of Hume's argument for the elusiveness thesis. Hume's first ploy is to deny that he has a proper idea of self on the grounds that his experiences are not of the right sort to generate the idea of an unchanging enduring self. But that assertion—the assertion that he has no clear idea of the self—undermines the second part of his argument. If he has no clear idea of the self then, as Shoemaker points out, it is not clear how he can be certain that he has not chanced upon it after all. If one has no clear idea of what a teakettle is, then one will not be in a position to judge whether one currently observes a teakettle. Without a clear idea of the self, we cannot

know whether one is available to introspective observation or not.

We might respond to Shoemaker, though, by pointing out that what it is that Hume wants to emphasize is that we have no idea of self on the basis of our perceptual experience, that is, that the idea of self we have is not a simple idea, generated by a sense impression of the self. Rather, the idea of the self that we have might be thought to be generated by the faculty of imagination, the faculty that Hume says is responsible for our mistaken notion that our minds exhibit identity through time. Hume knows very well what he is looking for, but he submits that the idea of self we operate with is one that is not 'clear and intelligible', because it is not based on perceptual experience.

There are several other aspects of the elusiveness thesis that might arouse criticism. Roderick Chisholm (1969) points out that each of us is only able to maintain that one individual—oneself—fails to introspectively observe a self. Hume admits that it is for each individual to confirm or disconfirm his claims for themselves (T 1.4.6.3, p. 165). The elusiveness thesis is based on Hume's claim that *he* introspects only particular experiences and thoughts, and no self. Each one of us must substantiate Hume's elusiveness thesis for ourselves. Moreover, there is a contingency to this version of the elusiveness thesis. Given that it is an empirical claim, it can be no more than that, as a matter of fact, when he has introspected, Hume has observed no self (Bermúdez, 2003, p. 215). And each of us, when we perform the same introspective act, can find that individually, only as a matter of fact, we observe no self. We should, therefore adapt the elusiveness thesis:

Elusiveness thesis (2): *When I introspect and thereby find out about my mental states, I do not observe any self to whom those states belong.*

Putting the thesis in the first-person in this way brings out a further criticism that is levelled at Hume's claims about the self. Chisholm (1969) objects to Hume's ontological conclusions concerning the nature of the self on the basis of a flaw he diagnoses in Hume's reasoning. Hume's evidence for his claim that he only observes experiences and thoughts when he introspects seems to show, according to Chisholm, that he encounters rather more than that: the self that Hume claims not to find is the self that uncovers those experiences and thoughts. The self that Hume looks for is the self that does the stumbling on those perceptions. The problem Chisholm has with Hume's claims is that they are, and must be, presented in the first person: '*I* never catch myself at any time without a perception'. As such, Chisholm takes Hume's evidence as implying not just the existence of those experiences and thoughts, but that there is a subject of experience also (Chisholm, 1969, p. 10).

Moreover, Chisholm contends that Hume is unable to express the same point in non-first-person terms such as 'Nothing but perceptions are found'. If he were to do this, his claim could no longer be classed as empirical (Chisholm, 1969, p. 11). If Hume were to make the general claim that only thoughts and experiences could be introspected he would be stepping beyond the purview of his own experience. He would be asserting that no one has ever introspectively observed anything like a self—a claim which would be unjustified by Hume's own experiences alone. So Chisholm suggests that Hume finds not only his thoughts and experiences when he introspects, but finds also *that* he finds his thoughts and experiences and according to Chisholm, it is from this that we can derive the idea of self.

Is Chisholm's objection a problem for the elusiveness thesis? His objection can be broken down into two parts. On the one hand Hume, as an empiricist, is duty bound to report the results of his introspective search for a self in the first-person because he can provide as evidence only his own experience. On the other, Hume reports that *he* stumbles on his thoughts and experiences. Chisholm adduces from these two facts that Hume is, in fact, aware of a self, a self that stumbles on thoughts and experiences. The first-person nature of the report of stumbling on thoughts and experiences is taken by Chisholm to imply the existence of a stumbling self. But in claiming that there must be a subject of experience if there is an experience of stumbling upon experiences and thoughts, Chisholm does not offer a rebuttal of Hume's claim that he does not introspectively *observe* a self. Hume is not licensed to conclude that no self exists on the basis of his own empirical investigation, but the elusiveness thesis (2) still holds. What is more, Chisholm's claims reveal why it is that Hume's elusiveness thesis generates a problem for the introspectionist.

Introspectionism is the view that the content of first-person thought is to be explained by the introspective awareness we have of the subject of thought and experience—the self. We saw in the previous chapter that the guaranteed reference of the first-person indicates that, if this is correct and we do employ an introspection-based self-concept when we entertain 'I'-thoughts, then that self must be available to introspective awareness whenever we are able to entertain an 'I'-thought. All and any 'I'-thoughts will refer successfully, so if they refer to the subject of thought and experience, that subject of thought and experience must be available to introspection whenever the subject is able to entertain an 'I'-thought. Hume reports the findings of his introspective exercise of searching for a self in the first-person: he expresses the first-person thought that he finds no self amongst the objects of introspection. So, whilst being able to entertain an 'I'-thought, he is unable to observe a self amongst the contents of his mind. At least in Hume's case it cannot be that the content of the self-concept

is introspection-based. Hume invites each of us to perform the same introspective enquiry; anyone who, like Hume, finds no self will add to the evidence that tells against the introspectionist. Moreover, even if someone were to announce that they have found an enduring self amongst the objects of introspection, this would neither indicate that introspectionism is correct, nor override any of the individual claims about the elusiveness of the self to introspective observation. As long as someone fails to introspect a self whilst being able to entertain first-person thoughts, then it cannot be the case that the content of the self-concept is to be explained in terms of introspection of a self.

If we do not introspectively observe or become acquainted with a subject of thought and experience, then it seems to be something of a mystery as to how introspection can provide us with the mode of presentation of the subject of thought allowing us to think about ourselves. Even if some people are able to observe a self introspectively, the absence of a general capacity for self-observation upon introspection amongst those that can entertain 'I'-thoughts leaves introspectionism with serious problems. If I can entertain 'I'-thoughts even though I do not introspectively observe a self, then it cannot be the case that there is an introspective mode of presentation of myself that gives content to the self-concept.

There is, however, another possibility that Hume does not consider: perhaps we can momentarily observe a self when we introspect. Hume fails to observe a persisting, substantial self, a self distinct from his experiences and thoughts that could bind those thoughts together over time. But, what of the possibility that we can observe a self when we introspect, even though we are not aware of that self as enduring? This kind of self-observation would not serve the purposes that Hume has in mind for a putatively experienced self: namely, providing evidence of diachronic identity. But it would provide us with a way in which the subject could be introspectively observed. It could be maintained that we introspect a subject of thought and experience only momentarily when we introspect. Holding such a view would be compatible with both a view of the self as enduring, and of the 'self' as a series of momentary selves (G. Strawson, 1997). If we are to introspect a self momentarily then it might be that the self is just a momentarily existing entity—each perception is a perception of a distinct self that exists only for that moment—or it might be that the self endures but that we can only perceive the self in perceptual experiences that last for only a moment. On this second view I only perceive what is in fact an enduring self for a moment at a time. Might the introspectionist make use of momentary introspective observations of a subject of thought to defend her view against Hume's elusiveness thesis?

Of course, the suggestion is vulnerable to Hume's line of attack: if each individual

cannot observe a self even momentarily when she introspects, then she will be able to assert the following:

Elusiveness thesis (3): *When I introspect and thereby find out about one of my mental states, I do not observe any self, even momentarily, to whom that state belongs.*

What is more, another formulation of the elusiveness thesis suggests that subjects of thought and experience are not the kinds of things that can be observed, even for a moment.

2.4 Kant et al.: we *cannot* observe a self

According to the alternative version, the self eludes us because it is not *possible* for the self to be the object of observational awareness. The idea, simply, is that subjects of thought and experience are not the kinds of things that can be observed *qua* subjects. Here I am taking the subject to be the subject of thought and experience, and the object to be the object of thought and experience—the object of attention.

Elusiveness thesis (4): *When one introspects and thereby finds out about one's mental states, one cannot observe the subject to which those states belong.*

The idea that there is something necessarily elusive about the subject can be found in the work of many philosophers, including Schopenhauer (1969), Wittgenstein (2001) and Sartre (1960), amongst others. But there are, in fact, two distinct ways of construing this version of the elusiveness thesis. According to some, the self is 'systematically' elusive, but not secure from objectification: it is only the subject of one's *current* experience that cannot be an object to itself. For others the self *qua* subject can never be satisfactorily objectified: when we think about ourselves, and if we were to observe a self, then what would be observed or thought about would not be the self *qua* subject of thought and experience but, to use Robert J. Howell's expression, "its proxy" (Howell, 2010, p. 462). I will suggest that, although the first view may also be correct, it is the second that provides us with the real problem concerning introspective observation of the self.

The view that there is something systematic about the elusiveness of the subject finds voice in Gilbert Ryle's (1949) description of the child looking for the thing to which 'I' refers:

"[…] the more the child tries to put his finger on what 'I' stands for, the less does he succeeds in doing so. He can catch only its coat-tails; it itself is always and obdurately a pace

ahead of its coat-tails. Like the shadow of one's own head, it will not wait to be jumped on. And yet it is never very far ahead; indeed sometimes it seems not to be ahead of the pursuer at all. It evades capture by lodging itself inside the very muscles of the pursuer. It is too near even to be within arm's reach." (Ryle, 1949, p. 186)

"Even if the person is, for special speculative purposes, momentarily concentrating on the Problem of the Self, he has failed and knows that he has failed to catch more than the flying coat-tails of that which he was pursuing. His quarry was the hunter." (Ryle, 1949, p. 198)

But, although the subject of one's thought remains always a step ahead of one's introspective observation, Ryle maintains that there is nothing "mysterious or occult" (Ryle, 1949, p. 198) about the subject of thought. According to Ryle there is nothing that is *essentially* elusive. The idea, as C.O. Evans (1970) points out, is that one's attempts to think about or observe oneself as the subject of thought and experience merely initiates an infinite regress:

"To hope to have an experience of the experiencer, if it makes sense at all, sets going an infinite regress. No matter how many manifestations of the self are experienced, there will always be one left over which has not yet been experienced, and that one is the subject of the last experience. Experience of the self, is, in Ryle's words, "logically condemned to eternal penultimacy"." (Evans, 1970, p. 148)

It is this infinite regress, the impossibility of providing a complete self-commentary that frustrates us. I can never provide that complete self-commentary because the act of self-commentary always generates one further fact to include in my self-description. We are therefore always 'on the coat-tails' of a complete self-description. It is because, in thinking about ourselves, we generate new facts about ourselves that we seem always to be hunting down something elusive. Yet, what eludes us at one moment, becomes the object of attention or awareness the next. There is nothing intangible about the subject of thought.

Does Ryle succeed in explaining away the seeming elusiveness of the subject? I think there is something right and something wrong about Ryle's elucidation of the problem. He is right to say that there is something systematic about the way in which the self eludes introspection. Certainly, the subject of a particular thought cannot take itself as an object. Yet, it is a mistake to think that there is no further mystery concerning the subject of thought and experience.

I.T. Ramsey (1955) criticizes Ryle's construal of the elusiveness thesis—and more importantly his dismissal of the problem—on the grounds that Ryle is not justified to assume that the objectification of the subject will leave the subject unchanged (Ramsey, 1955, p. 196). Ramsey's objection is that thinking about oneself performing a mental action will involve an objectification of the subject. One cannot take the subject of thought and

experience as the object of one's attention and thereby think about the subject of thought and experience *qua* subject.

This brings us to the second analysis of the elusiveness thesis (4). This second version of the thesis suggests that there is something *essentially* unknowable about the subject of thought and experience. We cannot dismiss the elusiveness of the subject of thought and experience as just the elusiveness of the current subject of thought from its own view, as Ryle does. Rather, the problem is that it seems that subjects are not the kind of thing that *can* be objects of experience and thought. Any attempt to introspectively observe the subject will not give rise to an observation of the subject, but only a copy or representation of the subject. As Howell characterizes the view, "[t]he self is elusive because all it yields to reflection is its proxy, not itself" (Howell, 2010, p. 462). He continues:

"it is part of the nature of the self to be a subject of perception or awareness, and something that is essentially a subject cannot be an object. At the very least, something crucial about the self is lost in its objectification—namely, the aspect in virtue of which it can truly be called a subject." (Howell, 2010, p. 463)

What eludes introspective observation is the subject of thought and experience *qua* subject of thought and experience. If we attempt to experience, or even think, about the subject, it will be the object of our awareness. As such it will no longer be a subject.

We might think of this idea as Kantian in spirit.[9] For Immanuel Kant the self cannot be observed, or intuited in Kant's terminology, by inner sense or by apperception. Kant's transcendental idealism ensures that we will never have knowledge of the noumenal subject of thought and experience, the subject as it is in itself, independent of the mind. On the one hand, the appearances of inner sense are conditioned by time, and therefore the objects of inner sense are never things in themselves, but mere appearances. On the other hand, the transcendental subject of apperception—the source of unification of the states of the mind—is not something that can be known at all.

First, Kant's transcendental idealism extends to self-knowledge: the same conditions apply to the objects of inner sense as apply to outer sense:

"Even the inner and sensible intuition of our mind (as object of consciousness) which is represented as being determined by the succession of different states in time, is not the self proper, as it exists in itself—that is, is not the transcendental subject—but only an appearance that has been given to the sensibility of this, to us unknown, being. This inner appearance cannot be admitted to exist in any such manner in and by itself; for it is conditioned by time, and time cannot be a determination of a thing in itself." (Kant, 1929, A492, p. 440)

[9] Again, I am not concerned with exegesis and offer no argument for the interpretation of Kant's *Critique of Pure Reason* that I present.

Thus, the things that we are aware of when we introspect are not the 'things in themselves', but only appearances conditioned by time. What is more, the transcendental unity of apperception, that which unifies the states of the mind, does not generate knowledge of the subject of thought (A350, p. 334). The subject of thought is not something that can be known, and any attempt to introspectively observe the subject will alter it:

"Through this I or he or it (the thing) which thinks, nothing further is represented than a transcendental subject of the thoughts = *X*. It is known only through the thoughts which are its predicates, and of it, apart from them, we cannot have any concept whatsoever, but can only revolve in a perpetual circle, since any judgement upon it has always already made use of its representation." (B404, p. 331)

When we attempt to think about ourselves as subjects of thought and experience we cannot take the pure subject as an object. Rather, we objectify it: the self *qua* subject becomes the self *qua* object. As Henry Allison (1983) puts it, "the subject of apperception, or more simply, consciousness, cannot grasp itself as an object because, first, in the endeavour to do so it succeeds only in negating its character as a subject" (Allison, 1983, p. 292).

So, the Kantian idea is that the self, understood as a subject of thought or experience, is logically elusive. Inner sense yields only appearance, and never the thing in itself. The transcendental unity of apperception yields nothing that can be classed as knowledge. And any attempt to introspectively observe, or even to think about the subject will objectify the subject, such that it will cease to be a subject.

I have considered two versions of the elusiveness thesis (4). The first version specified that the subject is logically elusive in the sense that the subject of a particular thought cannot be its own object. The second version claims that the subject is more comprehensively elusive. Any introspective observation of a self-like thing will not be an observation of the subject. Subjects, when they are taken as the objects of thought—*if* they can be taken as the objects of thought—cease to be subjects.

But if we cannot actually introspectively observe the subject of thought, then it cannot be the observation of that subject of thought that provides the content of the self-concept. Not only is it the case that some people profess themselves unable to observe a subject of thought when they introspect, even though they are able to entertain first-person thoughts, it is also the case that we can never actually introspectively observe a subject of thought *qua* subject of thought. It would have to be the case that we introspectively observe an 'objectified' subject, something that is no longer a subject of thought, or is a mere representation of the subject of thought, and it is this perception that comprises the content of the self-concept.

There is yet another reason, though, to object to the introspectionist's claim.

2.5 First-person thought and perceptual-demonstrative thoughts

As we have already seen, introspection gives rise to first-person thoughts that are identification free (Evans, 1982), and this has led us to model first-person thought on perceptual-demonstrative thought. But there are reasons for thinking that this is the wrong way to think of first-person thought.

The suggestion is that we introspectively observe a self and we are able to entertain first-person thoughts by individuating that self demonstratively. There are reasons, however, to reject the idea that first-person thought is a kind of demonstrative thought, and therefore reasons for rejecting the idea that the content of the self-concept is based in introspective observation.

The first significant difference between first-person thought and demonstrative thought is that the identification-freedom of first-person thought seems to be preserved in memory (Shoemaker, 1986, p. 111). By contrast, the identification-freedom of current demonstrative thoughts does not seem to transfer to demonstrative thoughts about the past.

I can think about myself in the past on the basis of my memory of the properties I instantiated. For example, I can think that I saw the Eiffel Tower on the basis of my memory of a perceptual experience that I previously had—my perceptual experience of the Eiffel Tower. My original present-tense introspection-based first-person thought is identification-free. My memory-based first-person thought about what I experienced in the past will also be identification-free. But the same cannot be said for demonstrative thought. A thought that *this* is red, whilst attending to a tomato, will be identification-free. However, if at a later time I point to a tomato that I then perceive, and think that *this was* red, then my judgement will be identification-dependent. My demonstrative thought about how things were in the past with the thing that I now perceive can be analyzed into two thoughts, one of which will be an identity judgement:

> "It goes with this that the past-tense demonstrative judgement rests on an observationally based reidentification of the thing referred to with "this," whereas the past-tense, first-person judgement does not rest on an observationally based reidentification of the person referred to with "I." "This was red" might be grounded in part on an observed similarity between the thing one sees now and the thing one remembers seeing to be red in the past, or it might be grounded in part on a series of phenomena one observed in perceptually tracking an object over time. "I was angry," if said on the basis of memory in the ordinary way, could not be grounded either on an introspectively observed resemblance between a past self and a

present self, or on an introspective tracking of a self over time." (Shoemaker, 1986, p. 111)

The point that Shoemaker wants to make is that self-ascriptions of introspected properties that are based on memory are identification-free.[10] Demonstrative thoughts that are based on memory are identification-dependent. I can only go from the present tense 'This is red' to the past tense 'This was red' if I know that the thing that I point to later on is the same thing that I pointed to earlier when I made my initial demonstrative judgement. I must re-identify the object.

Current observation of an object allows one to think about it demonstratively, entertaining an identification-free thought about the state of that object now. As soon as we attempt to think about an object over time, if we know about that object observationally and think about it demonstratively, then entertaining that thought will involve an identity judgement of some sort. The same can be said about demonstrative thoughts about our own bodies. For me to think about the state of the arm that I now see by my side—say, that my arm was broken last year—I must identify the object I perceive now (*this* arm) with the object I perceived last year (*that* arm). The fact that memory-based first-person thoughts are identification-free indicates that they do not involve any identity judgement between a currently perceived self and a previously perceived self. Shoemaker uses this distinction to suggest that first-person thoughts are not a type of perceptual-demonstrative thought in the sense that the individuation of oneself in first-person thought is not to be understood as based on introspective observation of a self.

There is a second reason that Shoemaker offers for distinguishing first-person thought from demonstrative thought: first-person thought has a particular role to play in the explanation of behaviour that usual kinds of demonstrative thought cannot play:

[10] It might be thought that the possibility of experiencing a quasi-memory (Shoemaker, 1970)—indistinguishable 'from the inside' from ordinary memory, but of a past event or experience that is not something that oneself underwent or enjoyed—should render our memory-based first-person thoughts identification-dependent. The possibility of quasi-memory brings out the fact that immunity to error through misidentification should not be identified with identification-freedom in all cases. In the usual case the identification-freedom of a thought ensures that the subject will be immune to error through misidentification. But in some cases the two might come apart (Bermúdez, 1998, p. 7). The possibility of quasi-memory ensures that memory-based past-tense first-person thoughts will be susceptible to error through misidentification. However that fact does not indicate that memory-based past-tense first-person thoughts are identification-dependent. Memory-based first-person thoughts are identification-free, yet when we entertain a thought that is in fact based on a quasi-memory our thought will be subject to error through misidentification. Shoemaker's point is that whilst past-tense memory-based first-person thoughts will be susceptible to error through misidentification, given the possibility of quasi-memory, still "first-person memory judgements do not involve identifications of oneself that are grounded on observed similarities between selves observed at different times, or on a perceptual tracking of a self over time" (Shoemaker, 1986, p. 111). The same cannot be said for past-tense demonstrative thoughts.

"To use David Kaplan's example, if I merely believe of the person I in fact am that his pants are on fire (I see someone in a mirror with his pants on fire, but do not realize that it is me), this will not influence my behavior in the way that the belief I would express by saying "My pants are on fire!" would." (Shoemaker, 1986, p. 112)

Shoemaker's worry is that a demonstrative account of the first-person, an account that subsumes the first-person within the class of demonstrative terms, would leave "totally unexplained the role of "I"-beliefs in the determination of behavior" (Shoemaker, 1986, p. 112). The role of first-person thought in the explanation of behaviour is one of the aspects of the cognitive significance of such thought that we looked at in Chapter One. First-person thought cannot be reduced to perception-based demonstrative thought because the latter kind of thought fails to have the cognitive significance of the former. We saw that this is the case for demonstrative thoughts based on ordinary kinds of perception already. Shoemaker argues that the same would hold true for thoughts in which we demonstrate an introspectively observed self. If the first-person element of first-person thought is best represented by the words 'this self', then we are going to have a problem explaining why first-person thoughts mean what they do to us. Why should I be particularly concerned about *this self* that I introspectively observe? We cannot think of the first-person as synonymous with 'this self' because if I believe only that 'this self's' trousers are on fire I will not be motivated to act in the way I would if I know that *my* trousers are on fire.

Shoemaker claims that first-person thought cannot be assimilated to perceptual-demonstrative thought because first-person thoughts about one's past are identification-free and because demonstrative thought fails to have the motivational force and cognitive significance of first-person thought.

2.6 The no-reference view

The introspectionist takes the content of first-person thought to comprise, in part, a perceptual mode of presentation of the subject of thought and experience. It is by introspectively observing the self that we come by this content. The identification-freedom of both first-person thoughts and perceptual-demonstrative thoughts motivates us to model the former on the later.

Since we can imagine being able to entertain 'I'-thoughts without being party to any bodily experience whatsoever, it seems that the resources for first-person thought must come from within thought and experience themselves. The suggestion is that when one introspects one becomes aware not only of one's mental states, but also the subject of those states, and it

is to this subject that one refers by employing an introspective observational—a perceptual—mode of presentation of that subject in first-person thought.

Introspectionism seems able to account for the fact that first-person thought is self-conscious thought, and that it has guaranteed reference. In particular, the guaranteed referential success of the first-person seems to entail that the object which one refers to must always be present when one is able to entertain a first-person thought. In the last three sections I have been examining why this cannot be the right account of the content of the self-concept. Not only do we have the reports of individuals who say that they do not observe a self or subject of thought, but we also have philosophical reasons for thinking that subjects of thought are not the kinds of things that can be observed. The introspectionist theory is really a theory that assimilates first-person thought into the class of perceptual-demonstrative thoughts. One refers to an object that one currently perceives by referring to it demonstratively: this is achieved by the employment of a perceptual mode of presentation of the object in question in the content of thought. But we have seen there are several reasons for thinking that first-person thoughts are not just a kind of demonstrative thought. The identification-freedom of memory-based first-person thoughts distinguishes first-person thought from perceptual-demonstrative thought since no such identification-freedom obtains for memory-based demonstrative thoughts. Even though introspection of a self would provide one with a way of thinking of oneself, such a thought would lack the motivational force and cognitive significance of first-person thought. While introspective observation of a self would give one a way of thinking about oneself, it is not a way of thinking about oneself as oneself. It is this latter requirement that a thought must meet if it is to count as genuinely self-conscious thought. Taken together these objections tell against introspectionism giving the correct account of the content of the first-person concept.

Introspectionism seemed to be the only possible way of giving content to the self-concept because we seem to be able to entertain 'I'-thoughts when we have no awareness of anything but that which we introspect. So if it fails we seem to have to conclude that first-person thought is non-referential. This in itself is a problem, but it also leads to another problem: the conclusion that the first-person pronoun is a non-referential term.

According to a Fregean semantic theory, a referring expression in language will express a sense, which will be the contribution of the expression to the thought expressed by the sentence in which it appears. It is the sense expressed by a term that determines its semantic value. Singular terms will express senses that determine the referent of the term. So, the first-person pronoun expresses a sense (the self-concept) that determines the reference of the

term. If there is no sense, then there will be no reference.

We have good reason for thinking that the claim that the first-person pronoun is not a referring term is false. Not only does the first-person pronoun have the same syntactic place in sentences as a referring expression ('I am tired', 'Alisa is tired'), but tokens of 'I' in a sentence uttered by *x* can be replaced *salva veritate* by the name '*x*' (Garrett, 1998, pp. 106-7). As Shoemaker says:

"The statement "Jones is thinking," said by someone other than Jones, is true when and only when Jones could truly say "I am thinking." […] So when someone says "Jones is thinking," and Jones says "I am thinking," both appear to be saying the same thing, namely that Jones is thinking. But there is no doubt that the word "Jones" in "Jones is thinking," refers to something. It refers to Jones. Since this is so, and since the statement "I am thinking" said by Jones, says the same thing, states the same fact, as the statement "Jones is thinking," said about him, it appears that the word "I", as used by Jones, must also refer to something;" (Shoemaker, 1963, p. 12)

"The idea that the word "I" stands for or refers to something […] has its basis, […] in the fact that corresponding to any first-person statement there are third-person statements that are in a certain sense equivalent to it and are certainly "about something."" (Shoemaker, 1963, pp. 12-3)

We need to offer some alternative account of how it is that the first-person pronoun is a referring term: either by offering an alternative account of the sense of the term—finding some other way of characterizing the content of the self-concept—or by offering an alternative to the Fregean picture of reference. In the next chapter we will see how reference might be fixed for the first-person pronoun and the self-concept. By appealing to a reference-rule that fixes the reference of the term and gives the universal meaning of a term that refers to different individuals in different contexts, we can save the linguistic term and concept from being classed as non-referential.

2.7 Conclusion

To summarize, we have examined a number of different claims concerning the elusiveness of the self. In Hume's *Treatise* we found the claim that when we introspect we do not observe a persisting, substantial self. In a more Kantian vein, we also considered the suggestion that subjects of thought and experience are not the kinds of things that can be observed. We examined the claim that first-person thoughts about oneself in the present and in the past are identification-free, and the ramifications of this claim for the view that one thinks about oneself on the basis of introspective observation. Lastly, we noted that

perceptual-demonstrative thought does not have the cognitive significance of first-person thought.

Of course, we might seek to address these concerns by simply indicating the observability of the body. One can observe one's own body, and perhaps such observation should be considered to be observation of the self or subject of thought and experience. But the matter is not dealt with so easily: the elusiveness thesis in no way denies that bodies, even 'subject-bodies' can be observed. The problem is that through introspection alone one becomes aware only of one's thoughts and experiences. The claim is a problem because it seems that the guaranteed referential success in combination with the fact that first-person thought is self-conscious thought pushes us towards the view that on the basis of introspection alone one is able to entertain first-person thoughts. One's first-person thoughts essentially concern oneself in such a way that one must be present to oneself whenever one is able to entertain a first-person thought. One is able to entertain first-person thoughts on the basis of introspection alone, so that whatever it is that one must be related to in order to entertain a first-person thought, that object will be present when one introspects. And we have seen that no such object seems to be available to introspective awareness.

We are therefore left with a problem in explaining the content of the self-concept: it seems that introspection is the only possible means by which we could come by the right kind of content, but in fact it presents us with no awareness of ourselves. The self-concept does not seem to be a referring tool after all, and this also puts the referentiality of the first-person pronoun at risk as long as we take the contents of thought to determine the reference of the linguistic terms that are used to express those contents.

Chapter 3: The self-reference rule and self-consciousness

3.1 Introduction

In the previous chapter we saw how modelling first-person reference on demonstrative reference, such that the content of the self-concept expressed by a use of 'I' is characterized in terms of the introspective observational information available to the subject, leads us into problems. If we expect perceptual or quasi-perceptual information available to the subject when she introspects to comprise the content of first-person thought, and thus to determine the reference of utterances of sentences that express those thoughts, then we fail to provide an account of first-person thought or first-person reference. The self or subject of thought eludes our observation when we introspect, and hence it seems that the first-person does not refer.

But the conclusion that the first-person pronoun does not refer is unacceptable, not least because it would mean that we do not express truth-evaluable propositions when we speak about ourselves using the first-person pronoun. We need to reconsider our model of reference. In this chapter I will look at a different model for reference of the first-person pronoun, and see what it can tell us about first-person thought. On this new (to us) approach, the reference of the first-person pronoun is not taken to be determined by one's perceptual experience of oneself, but rather reference is fixed by a rule. Not only can this view, which I shall call the self-reference rule account (O'Brien, 2007, p. 49) be used to defend the referentiality of the first-person pronoun, it can also accommodate the immunity from reference failure and immunity from mistaken reference that seem to mark the first-person as special.

Whilst it has these clear advantages, the view has had its detractors. Anscombe, in addition to arguing against the perceptual demonstrative model of first-person reference, objects to the idea that the self-reference rule "explains what 'I' names, or explains 'I' as a "referring expression" " (Anscombe, 1994, p. 142). Her criticism of the view is that it fails to explain the self-consciousness that first person pronoun-use expresses. Avoiding this inadequacy, she says, results in circularity. But, whilst Anscombe's argument indicates that we may need to consider how it is that speakers can come to use a term governed by a reflexive rule (O'Brien, 1994), she fails to show the inadequacy of the self-reference rule account as an account of first-person reference.

Whilst Anscombe is wrong to think that the terms in which the self-reference rule is specified must account for the fact that we use the first-person pronoun to express self-consciousness, her argument indicates that we still need to worry about the content of first-person thought. Even if the self-reference rule account secures the referentiality of the first-person pronoun, it fails to provide us with an account that respects the fact that first-person pronoun use expresses self-conscious thought: thought in which a speaker knows that she is referring to herself.

3.2 Indexicality and the self-reference rule

The conclusion reached in Chapter Two that the first-person pronoun does not refer depends on the assumption that in order to refer to an object one must stand in some kind of epistemic relation with that object. According to this view demonstratives are the paradigm referring expressions, and the reference relation that obtains between language and world, where it does obtain, is dependent on the subject's experience of the world. This is just a relaxation of Russell's view of singular thought: rather than requiring an epistemically narrow relation of acquaintance to obtain between referrer and referent, with the unattractive consequence that we can only entertain singular thoughts about sense data, we allow that ordinary perception of objects in the world provides the right kind of epistemic connection between thinker and world for a subject to refer to objects in the world.

The argument for the no-reference view also turns on marrying this view of singular thought with a Fregean semantic theory, according to which referring terms express a sense, and it is this sense that determines the reference of the term. The reference or semantic value of a term is the contribution that the term makes to the truth-conditions of the sentence in which it appears. For a singular referring expression the reference will be the object that is the referent of the term. But, according to the Fregean, this is not all that must be said about meaning. Distinct form the semantic value or the reference of a linguistic term is the thought, or sense, expressed by the sentence in which the term appears, and, as a component of that, the sense of the linguistic term itself. The Fregean complete sense—the sense of a sentence—is equated with the content of thought or the proposition expressed by an utterance of the sentence. The sense of a singular term will be the mode or manner of presentation of the referent, that is, the way in which the thinker is thinking of the referent. For a demonstrative or quasi-demonstrative term the sense of that term will be based on the perceptual encounter that a subject has with the referent. One's perceptual experience

determines the object as the referent of the demonstrative one uses.[11] Sense provides a way of explaining both the cognitive significance of informative identities and the rationality of holding different epistemic attitudes to the thoughts expressed by co-referential sentences, as we already saw in Chapter One. And, for the Fregean, without sense we have no reference. It is for this reason that I argued in Chapter Two that the first-person pronoun does not appear to be a referential term. Since we do not introspectively observe any self, there seems to be no presentation of an object that could determine the reference of a use of the term, and hence, since sense is taken to determine reference, there is no reference.

The conclusion that the first-person pronoun is not a referential term is far from desirable, and calls for us to reconsider our view of first-person reference. What is more, we need to take into account a feature of the first-person pronoun that suggests that we should adopt a very different account of its reference: the first-person pronoun is an indexical term. Indexical terms are context-sensitive expressions. By this I mean that the reference of the term will vary from context to context. My utterance of 'I' does not have the same reference as your utterance of 'I': my utterance refers to me, whilst your utterance refers to you. Other indexical terms include 'today', 'here', and 'now'. Demonstratives such as 'this' and 'that' are also context-sensitive. Their reference is determined in part by the context of utterance. However, they are not what are known as 'pure indexicals' since their reference depends also on a demonstration of the object that is being referred to. Pure indexicals (which I will simply call indexicals) such as 'I' or 'now' require no demonstration.[12] So, the reference of an indexical term will depend on the context of utterance. That context might include the agent of the context—the speaker—as well as the time and place at which the utterance was produced, and possibly also the world at which it was produced.

Indexicals, whilst their reference will be different in different contexts, also have a shared or common meaning across contexts. A token of a particular type of an indexical has something in common with other tokens of that type. My utterance of 'I' has something in common with your utterance of 'I'. If we both produce an utterance of the sentence 'I am happy', then, even whilst my token of 'I' refers to me and your token of 'I' refers to you, there is some sense in which we have said the same thing.

Both the shared meaning that our utterances of the first-person pronoun seem to have, and the context-sensitivity of the reference of the first-person pronoun make the Fregean

[11] I take Gareth Evans (1982) to present this kind of view of reference, which takes Russell's notion of singular thought and adapts it to allow for singular referring terms to express Fregean senses. I will discuss the specifics of Evans's account of first-person thought in Chapter Five.
[12] Although see de Gaynesford (2006)

semantic approach a poor paradigm in which to think about the term. Fregean semantic theory assigns a sense to a term, and whilst several senses might attach to a single term, there is no provision for the sense of a term to differ in different contexts. And whilst tokens of the same term may express the same sense (for example, the sense expressed by my utterance of 'Barack Obama' might be the same as the sense expressed by your utterance of 'Barack Obama'), that sameness of sense—which could be read as the shared or common meaning of the term—does not accommodate differences in reference. For Frege, although the same reference can be determined by different senses, same sense will give the same reference in all contexts.[13]

We need an account that explains the indexicality of the first-person pronoun adequately. David Kaplan (1989a, 1989b) and John Perry (1977) give similar suggestions about how we should explain the context-sensitivity of an indexical term, each giving an account according to which the reference of an indexical term is determined by a rule.[14] The rule determines individuals or objects as the reference of a token of the term within the context of utterance. This rule—known as the character or role of the term—is the common or shared meaning of the term. Both your and my utterances of 'I' have a shared meaning, which is a rule particular to the first-person that determines reference. But the rule determines that reference within the context in which the utterance is produced. We can think of the rule as a function from context to individual or object. We apply the context in which an utterance has been produced to the rule and yield a semantic value or reference. This allows us to explain the context sensitivity of the term. If reference is determined by a common or shared rule in conjunction with the context, then the reference will depend on the context in which an utterance is produced. Across different contexts the first-person pronoun will have different references.

The rule for each indexical term must specify which aspect of the context of utterance is essential to the determination of the reference for that term. For the first-person pronoun it is the speaker that is the relevant feature of the context upon which the reference of the token of the term will depend. There are a number of ways of spelling out the rule that is supposed to determine the reference of the first-person pronoun. For example, John Campbell says, "The term *I* is governed by a simple rule: any token of *I* refers to whoever produced it" (Campbell, 1994, p. 73). Bermúdez gives the rule as the following: "When a person employs

[13] I do not intend for my discussion here to represent Frege's view of demonstratives, but rather merely to illustrate the inadequacies of a general Fregean semantic programme in accounting for demonstrative terms.

[14] Perry (1977) is primarily concerned with providing an account of demonstrative thought, whereas Kaplan's (1989a, 1989b) is a semantic account of demonstrative terms.

a token of 'I', in so doing he refers to himself" (Bermúdez, 1998, p. 14),[15] and O'Brien says that, "The self-reference rule states that 'I' refers to the subject who produced it" (O'Brien, 2007, p. 49). The general idea is that reference is fixed by a reflexive rule, a rule that specifies that the reference of an utterance of 'I' refers to the producer of that utterance. In using 'I' one is speaking about oneself.

I will use the following statement of the rule: a token utterance of 'I' refers to the person who produced it. We can call this rule the self-reference rule (O'Brien, 2007). The rule by itself doesn't determine a reference. It is in conjunction with the context in which the term is used that the rule determines or fixes the reference. My being the agent of the context, the producer of a particular utterance, fixes me as the reference of that utterance, because the rule specifies that it is the producer of a token utterance, or the agent of the context, that is of concern.

3.3 The self-reference rule and the content of thought

However, whilst we might find the idea that reference is determined by the self-reference rule together with the context in which an utterance is produced attractive since it can account for the context-sensitivity of first-person reference, we still haven't said how the rule is supposed to function. The self-reference rule is supposed to provide a way of determining the semantic value of an utterance of 'I', so should we equate the rule with the sense expressed by a term, and thus with the content expressed by the term? The suggestion is that we incorporate the rule into a Fregean semantic theory.[16]

Of course we can note a way in which the account of reference determination on offer differs from the Fregean approach. As Perry points out: "Senses do not carry us from context to references, but directly to references, the same on each occasion of use" (1977, p. 6). Sense, as the Fregean understands it, is not a function from contexts to references. But, this in itself might not dissuade us from the view that the content of thought is to be characterized by the self-reference rule. Perhaps the content of thought for other kinds of terms comprises the Fregean sense of the term, but for indexicals we need to characterize the content of thought differently—in terms of a rule that is a function from contexts to semantic values. By doing so we are able to accommodate the context-dependence of indexical

[15] Bermúdez considers several different formulations of the self-reference rule, but his discussion of these formulations pre-empts my discussion of Anscombe's objections to the self-reference rule account in the following sections.
[16] Compare with Frege's account of demonstratives.

reference.

In favour of this interpretation of the self-reference rule account of first-person reference, Kaplan and Perry both suggest that it is the character of the first-person pronoun—the self-reference rule—that explains its cognitive significance.

Recall that in Chapter One we saw that it seems rational for subjects to hold different epistemic attitudes to propositions expressed by sentences in which co-referential terms are employed, and that co-referential identity statements can be informative. To explain these phenomena Frege introduced the notion of sense as distinct from the reference or semantic value of a term. The sense of a term not only determines the semantic value of that term, but presents that semantic value to the subject. The sense is therefore a way of thinking of the semantic value of the term, and it is in this way that cognitive significance is explained.

According to Kaplan and Perry we explain cognitive significance in terms of the character of the indexical term used to express a proposition. So, does this mean that Kaplan and Perry equate the self-reference rule with the content of thought expressed by an utterance of a sentence in which the first-person pronoun occurs?

They do not, and the interpretation is mistaken: Kaplan and Perry have a different semantic approach from the Fregean one. For Frege, we can equate the aspect of meaning that explains cognitive significance with the content of thought, which is also equated with the proposition expressed by an utterance. Two sentences in which co-referential terms occur will have differing cognitive significances for a subject because they express distinct propositions or thoughts. There is a sense in which the Fregean adopts a two-dimensional semantics, since the content of thought expressed by an utterance of a sentence is to be distinguished from the truth-conditions of that sentence, which is the level or dimension at which we find the semantic value of a referring term.

Kaplan (1989a, 1989b) and Perry (1977) also advocate a two-dimensional semantics, but theirs is of a different kind to Frege's. They also demarcate the roles that different kinds of meaning play in a different way to Frege.

As direct reference theorists about indexicals, Kaplan and Perry take an indexical term to contribute its semantic value or reference to the propositional content expressed by an utterance of a sentence in which the indexical term occurs. The content expressed by the first-person pronoun is the individual herself to which the term refers. There is also another dimension of meaning for Kaplan and Perry, but it is distinct from the propositional content expressed by an utterance of a sentence. This is the conventional or linguistic significance of a term. A subject who uses or hears an utterance must know this conventional or linguistic

significance if she is to understand what is said. In the case of indexicals, the linguistic meaning is a rule that determines the contribution that the indexical makes to the proposition expressed within a context. For the first-person pronoun, the conventional meaning is given by the self-reference rule. The self-reference rule is not a part of that content, but is only the linguistic or conventional meaning of the term, providing a way of determining who it is that is part of the propositional content.

According to Kaplan and Perry, we explain cognitive significance not in terms of different propositional contents that are entertained by individuals as Frege does, but in terms of different ways of grasping one and the same proposition. The proposition expressed by an utterance of the sentence 'I am happy' by me, is exactly the same proposition expressed by an utterance of the sentence 'Alisa is happy' by me, or by someone else. Both utterances express the proposition:

{Alisa, the property of being happy}

As directly referential terms, both my utterance of 'I' and an utterance of 'Alisa' contribute me to the propositional content expressed by the utterances of the sentences in which they appear. But your utterance of the sentence 'Alisa is happy', and my utterance of the sentence 'I am happy' offer different ways of grasping that proposition. I grasp the proposition under the character of the first-person pronoun, whilst you grasp the proposition under the character of the name 'Alisa'. Far from giving us reason to equate the self-reference rule with part of the propositional content expressed by an utterance of a sentence in which the first-person pronoun is employed, the explanation of cognitive significance in terms of the self-reference rule is the result of Perry and Kaplan equating the thing that explains cognitive significance with the aspect of meaning that is external to the propositional content grasped by a speaker or a hearer.

So, according to Kaplan and Perry we should conclude that the self-reference rule determines the reference of an utterance of the first-person pronoun, but not by comprising the content of the proposition expressed by an utterance of a sentence in which 'I' is employed.

There is a more general reason, as well, for distinguishing the self-reference rule from the content of thought. We need not accept Kaplan and Perry's direct reference view of propositional content to think that if the self-reference rule determines reference, then it cannot be by comprising a part of the content of thought expressed by an utterance of the first-person pronoun.

Taking the content of thought to comprise the rule that a token utterance of 'I' refers to

the person who produced it would have the result that speakers who utter the same sentence types in which indexicals are employed would be entertaining the same propositions also (Perry, 1977, pp. 7-8). Suppose that you and I both say 'I am happy': you say 'I am happy' and I say 'I am happy'. And suppose that your utterance is true whilst mine is false: you are happy but I am not. Now, we both express a proposition that will have the same content with respect to the predicate phrase '…is happy'. And we both use the first-person pronoun in our utterance. So, it seems that you and I should be expressing the same proposition when we speak, if we take the self-reference rule to be the part of the content of thought expressed by an utterance of 'I'. Yet, your utterance expresses a true proposition whilst my utterance expresses a false proposition—they cannot be the same. If the content of the self-concept was characterized by the self-reference rule, then we could not account for the difference in the propositions we express, specifically in the difference in truth-values of the propositions expressed by different individuals. If the self-reference rule gives an account of first-person reference, it does so by fixing the reference but not by characterizing the content of thought.

3.4 Referential success

The self-reference rule account of first-person reference is attractive not only because it can account for the context-sensitivity of the term. It is also appealing because it requires so little, epistemically speaking, of the speaker. Recall that in Chapter One I detailed the special referential security that the first-person seems to have. Our uses of the first-person are guaranteed to refer and to refer to the individual intended by the speaker (Bar-On, 2004, p. 72). When we use the first-person pronoun we are guaranteed a referent, and that referent will be the one intended by the speaker. These immunities from failure appear to be at odds with any view that takes the reference of the first-person pronoun to be determined by perceptual experience, for perception can always be disrupted. Even when a speaker does not have any experience of her own body and the world, or she experiences some object rather than herself in the 'first-person' way, she is still able to use the first-person pronoun correctly to refer to herself. Guaranteed reference proves to be a problem for the perceptual demonstrative account of the first-person pronoun because that view accounts for the determination of reference in terms of the content of one's thought expressed by an utterance of the first-person pronoun, which itself comprises one's perceptual knowledge of oneself. The subject must stand in an epistemic relation to the object that is the referent of the term—she must have perceptual knowledge of herself. But, as we saw in Chapter One, perceptual

knowledge of ordinary objects like bodies can always fail, either because the subject fails to have any perceptual experience of her body, or because she experiences something else which she mistakes for her body.

The possibility that we receive perceptual misinformation or no perceptual information about ourselves renders our perceptual systems fallible. If, however, first-person reference is to be determined by perceptual information about ourselves, and the perceptual apparatus by which we come by this information are fallible, then first-person reference cannot be guaranteed to be successful. The problem with the demonstrative model, and any model according to which reference is determined by knowledge of the referent is that whatever method we use to come to know about the referent, that method can come unstuck. By contrast, the first-person pronoun seems to 'stick' to its referent in all and any circumstances (Bar-On, 2004, p. 76; DeVidi, 2001, p. 32).

The incompatibility of the guaranteed success in referring that we enjoy in our first-person pronoun use with the perceptual demonstrative model of reference leads Anscombe to argue that the first-person pronoun does not refer at all. She seems to assume that if 'I' is a referring term it must be a demonstrative, and hence the reference of the term will be dependent on our experience. The guaranteed success in referring that we enjoy suggests that whatever object is the referent of the term must always be present to the referring subject. Since, it is argued, we could continue to use the first-person pronoun even in circumstances in which we had no experience of our bodies whatsoever, it must be the case that the reference of the first-person pronoun is determined by our experience of a Cartesian ego. Anscombe reasons from the absurdity of this verdict to the conclusion that the first-person is not in the business of referring.

However from the premise that our uses of the first-person pronoun enjoy guaranteed referential success, we need not accept Anscombe's conclusion that the first-person pronoun does not refer. Rather, we might take it to show that a demonstrative model was the wrong one to use to account for first-person reference. The guaranteed referential success of our 'I'-use seems to present two options. Either we posit a "mysterious capacity for transparent self-perception" (O'Brien, 2007, p. 7), or we look for an account that does not place a perceptual epistemic burden on the subject in order to account for the determination of reference. Anscombe takes the first option, only to reject it as absurd and claim that the first-person is not a referential term. We might, in fact, take the absurdity of Anscombe's conclusion that 'I' is not a referring term as a *reductio ad absurdum* of the view that one can stand in the kind of perceptual relation to an object that is secure against failure to perceive or

misperception:

"It is in the nature of perception to allow for the possibility of a rational subject suffering from 'blindness' and mis-perception. The problem with attempting to fashion a perceptual model of first-person reference comes, in essence, from the fact that we take the subject who suffers such perceptual disruption to be able, nevertheless, to express herself, as herself, using the first person." (O'Brien, 2007, p. 48)

The alternative is to look for an account that does not place any unnecessary epistemic requirements on the subject in order to account for the determination of reference. According to the self-reference rule account of first-person reference, the rule in conjunction with the context is sufficient to fix the reference of the term. My utterance of 'I' is about me because the self-reference rule specifies that the reference of the token of that term will be the individual that produced that utterance, and in the context in question I produced the utterance. So, it is not the case that reference is determined by the perceptual information that a subject has about herself, which she employs in the first-person thought that she entertains when she produces a meaningful utterance of the first-person pronoun. The first-person pronoun will refer even if the subject has no perceptual information about herself, from the inside or out. The subject need have no perceptual knowledge of herself, either as a subject of thought or as a material body, and her utterances of 'I' will refer. The fact that the self-reference rule account of first-person reference places no perceptual epistemic demands on the subject means it can accommodate the guaranteed referential success that we enjoy when we use 'I'.[17]

In fact, referential success seems to be a consequence of the self-reference rule account of first-person reference. Since the rule specifies that any particular token of the first-person will refer to its producer, then there will be no token of 'I' that does not have a referent. The self-reference rule seems to make it a matter of necessity, albeit *de facto*, that every token of 'I' will have a referent. By dint of being produced at all, given the content of the self-reference rule, the term is guaranteed to have the subject as its referent:

"Anyone who is capable of formulating such [an 'I'] thought will have mastered the ordinary practice of personal reference by the use of personal pronouns; and it is a rule of that use, that the first-personal pronoun refers, on each occasion of its use, to whoever then uses it. So

[17] As we will see in the next sections, a comprehending utterance of the first-person pronoun by a speaker will require that a speaker uses the term *as* a term governed by the self-reference rule. Although this will not require that the subject have any specific formulation of the rule in mind, she must have some understanding that her utterance refers reflexively. So, the self-reference rule account of first-person reference places some epistemic demands on a speaker. However a subject's knowledge of the self-reference rule or its analogue is not a form of perceptual knowledge.
We will see, however, in Section 3.6 that there is a further demand we must place on a speaker if her utterances are to count as expressions of self-consciousness, and it is for this reason that the self-reference rule account of first-person reference fails as it stands as an account of our use of the first-person pronoun, or as an account of the content of first-person thought.

the fact that we have, in the case imagined, a user, is sufficient to guarantee a reference, and the correct reference for the use. It is not in the least necessary, in order for the guarantee to operate, that the user should know *who* he is." (P.F. Strawson, 1994, p. 210)

So far we have seen that an alternative account of first-person reference to the perceptual demonstrative model favoured in Chapter Two has it that the reference of 'I' is fixed by the self-reference rule plus the context of utterance. We have seen that this account can accommodate the special referential security that seems to attend our uses of the first-person pronoun, and we have also seen that it can account for the context-sensitivity of first-person reference.

The account presents us with two further questions. The first question concerns the adequacy of the view in accounting for the meaning of the first-person pronoun. I will discuss the claim that the self-reference rule account of first-person reference is inadequate in Section 3.5. The concern is that the self-reference rule will not give us an account of the reference of the first-person pronoun. In fact, the worry is unfounded: the self-reference rule in conjunction with the context determines the producer of a token of the first-person pronoun as the reference of that token. As a semantic account, the self-reference rule is fit for purpose.

The second question concerns whether or not we need to posit any further knowledge on the part of a subject other than knowledge of the self-reference rule if she is to be capable of producing utterances of 'I'. Uses of 'I' are expressions of self-consciousness in that one knows that one is referring to oneself reflexively. For a speaker to use the first-person pronoun as a term governed by the self-reference rule she must use the first-person pronoun with the intention of referring to herself. I will argue that she can do this only if she has the capacity for first-person thought.

3.5 The self-reference rule and semantic adequacy

Part of the appeal of the self-reference rule account is the simplicity of the rule that fixes the reference of the first-person pronoun. The rule is simply the following: a token utterance of 'I' refers to the person who produced it. Moreover, it not only accommodates the guaranteed referential success of the first-person pronoun, but can help us account for that guaranteed success.

Yet Anscombe (1994) objects to the self-reference rule account. Her concern with the account is that it will either prove to be insufficient or circular, a view that is prompted by

the idea that reflexive reference is not identical to first-person reference (O'Brien, 2007, p. 56). What Anscombe believes is that the fact that the first-person can be used to express self-conscious thought means that no self-reference rule-based account of first-person reference can be adequate unless the rule is specified in terms that render the account circular. She claims, on this basis, that the rule cannot give the reference or the meaning of the term (Anscombe, 1994, p. 142). But this is wrong: the self-reference rule is adequate to determine the reference of a use of the first-person pronoun within a context, and is therefore adequate to "explain 'I' as a referring expression" (Anscombe, 1994, p. 142).

Anscombe begins her argument by offering the following statement of the reflexive rule that fixes the reference of the first-person pronoun:

SRR_{Ans}: "['I'] is the word each one uses in speaking of himself" (1994, p. 142)

The problem with SRR_{Ans}, and with any other formulation of the self-reference rule, Anscombe says, is that it does not capture what is special about first-person reference. The first-person is not a term of mere reflexive reference, which is how the self-reference rule characterizes it. The first-person is used for self-conscious self-reference, and the rule only individuates self-reference, both accidental and intentional (Bermúdez, 1998, p. 16).

To see the difference between accidental and intentional self-reference recall Perry's (1979) example of following a trail of sugar around a supermarket looking for the offending shopper who has a torn and leaking packet of sugar in his shopping trolley, only to discover that it is in fact he, himself who has been creating the trail because the torn packet of sugar is in his own trolley.

If Perry, prior to realizing that it is himself who has the torn packet of sugar, says 'The person who has a torn packet of sugar in their trolley is making a mess', then he will refer to himself. He will be using the description 'the person who has a torn packet of sugar in their trolley' to refer reflexively. But in that situation Perry does not know that it is himself that is making the mess. He refers to himself, but he does so accidentally.

So too, using a proper name to refer to oneself is a form of reflexive reference, but it too can be accidental if one is suffering from amnesia and fails to realize that the individual to whom the name refers is oneself. Suppose that Perry is still in the supermarket, but is now suffering from amnesia—he cannot remember his own name. He finds a credit card on the floor with the name 'John Perry' printed on it. When Perry says 'John Perry has dropped his credit card', he refers to himself for it is his own credit card that he has found. But again, his self-reference is only accidental. What marks the difference is that in each situation, Perry does not know that he is referring reflexively. He refers quite properly when he uses the

description or the name, and in each instance he intends to refer to the individual who satisfies the description, or who is picked out by the name, but his reference is not intentionally reflexive. He does not know that he is speaking about himself.

The self-reference rule, whether we characterize it as Anscombe does, or in some other similar fashion, will fail to individuate intentionally reflexive reference from unintended reflexive reference. Anscombe diagnoses the problem as arising from the ambiguous locution 'speaking of himself' (O'Brien, 1994, p. 277). The only way to avoid the insufficiency of the rule will be to stipulate that the third-person pronoun 'himself' be interpreted as what Anscombe calls the indirect reflexive (Anscombe, 1994, p. 141). But the indirect reflexive, roughly analogous to Castañeda's (1966, 1969) quasi-indicator 'he*', is a term that can itself only be explained in terms of the first-person pronoun. What we mean when we say that an individual is talking about himself* is that he is thinking about himself in a way that he would naturally express by an utterance of the first-person pronoun.

By stating the rule so as to differentiate first-person reference from mere unintended reflexive reference, we have made our account of the first-person circular:

"[…] the explanation of the word 'I' as "the word each of us uses to speak of himself" is hardly an explanation!—At least, it is no explanation if that reflexive has in turn to be explained in terms of 'I'" (Anscombe, 1994, pp. 142-3)

Anscombe envisions the self-reference rule as something like the following:

SRR_{Ans*}: 'I' is the word each one uses in speaking of himself*

We have supplemented the original version of the self-reference rule so as to avoid the charge that it is insufficient to individuate first-person reference, but this supplementation comes at the cost of circularity (O'Brien, 2007, p. 8).

But perhaps, as O'Brien (1994) points out, it was not the reflexivity of the self-reference rule that should be expected to do all the work of individuating first-person reference from other forms of reflexive reference. After all, 'I' is a word that *every* person uses to speak of himself. When I use the proper name 'Alisa' or the description 'the second daughter of Mikhail and Susan' then I refer reflexively, but other people do not use those particular terms to refer reflexively. What distinguishes the first-person pronoun from other terms, it might appear, is its generality (O'Brien, 1994, p. 278). The response to Anscombe's argument might then be that first-person reference is distinct from accidental self-reference because it is a term that *everyone* uses for self-reference.

To bolster her cause and protect it from this response, Anscombe offers the following thought experiment, which she takes to reveal the insufficiencies of the self-reference rule (O'Brien, 1994, p. 278):

"Imagine a society in which everyone is labelled with two names. One appears on their backs and at the top of their chests, and these names, which their bearers cannot see, are various: '*B*' to '*Z*' let us say. The other, '*A*', is stamped on the inside of their wrists, and is the same for everyone. In making reports on people's actions everyone uses the names on their chests or backs if he can see these names or is used to seeing them. Everyone also learns to respond to utterance of the name on his own chest and back in the sort of way and circumstances in which we tend to respond to utterance of our names. Reports on one's own actions, which one gives straight off from observation, are made using the name on the wrist. Such reports are made, not on the basis of observation alone, but also on that of inference and testimony or other information. *B*, for example, derives conclusions expressed by sentences with '*A*' as subject, from other people's statements using '*B*' as subject." (Anscombe, 1994, pp. 143-4)

Ex hypothesi 'A'-users are not self-conscious. The point Anscombe wants to make is that the simple rule that 'I' refers to whoever produces the token in question, and that it is a term that each of us uses to refer in this way, does not capture what seems to be special about 'I'-use: self-consciousness. 'I' is used in self-conscious self-reference, whereas 'A', ostensibly governed by the same rule as 'I', is used for self-reference with the express absence of self-consciousness. Thus, the self-reference rule is deemed inadequate.

Of course, we might object that 'A'-users are self-conscious because they observe themselves, and any form of self-observation is a form of self-consciousness. We can even add that 'A'-users observe themselves in a unique way. Anscombe asks that we think of their 'A'-use as being dependent on their observing their wrists from a first-person perspective. It is only when an 'A'-user sees his wrist from the first-person perspective that he will use 'A' to speak of the individual that he observes. But I think we can, nevertheless, grant Anscombe that 'A'-users are not self-consciously referring reflexively in the way that 'I'-users are.

However, this distinction does not support Anscombe's conjecture that the self-reference rule fails to account for the reference of the first-person. Instead, it helps us to see that whilst Anscombe is right to think that our use of the first-person pronoun is special, explaining why this is so is not required of a semantic account, but is the province of an account of the pragmatics of first-person pronoun use (O'Brien, 1994, p. 280; Bermúdez, 1998, p. 17).

In Anscombe's 'A'-using community, the reference of 'A' is not fixed by a reflexive rule, but is fixed instead by observation (O'Brien, 1994, pp. 278-9; McDowell, 1998, pp. 137-8). Presumably, given her intended use of the example, Anscombe does not interpret things in this way. She takes the reference of 'A' to be determined or fixed by the self-reference rule. But, as she describes the community, this cannot be right. An individual uses 'A' because it is the letter they see on a particular body, just as they use the letters 'B' to 'Z' to refer to individuals they observe when they observe one of those letters on a particular body (O'Brien, 2007, p. 57).

O'Brien's (1994, 2007) idea that use of a term can be *governed* by a rule helps us to characterize the difference between 'I'-use and 'A'-use. It is true that 'A'-use accords or is consistent with a the self-reference rule—that is, a producer of a token of 'A' will be the referent of that token—however, it is not *governed* by the self-reference rule. When the individuals in the 'A'-using community see the letter 'A' on an individual's wrist, they use the letter to talk about that individual. It just so happens that they tend only to see their own wrists, and so only use 'A' to speak about themselves. As such, their 'A'-use will be (more or less) guaranteed to be self-referential. But as Anscombe herself wants to stress, in addition to being guaranteed to be self-referential, first-person reference is intentional self-reference. It is because we use 'I' as a device of reflexive reference, with the intention of referring to ourselves, that it is special and distinct from other forms of reflexive reference.

O'Brien's distinction between reference in accordance with a rule and reference governed by a rule allows us to see where Anscombe goes wrong. The distinction amounts to a difference in the ways in which we come to refer: what is required for us to refer using a particular term. But this is a matter of what O'Brien and Bermúdez call the pragmatics of first-person use, rather than the semantics of the first-person pronoun. It is not the self-reference rule that must individuate intentional from unintended reflexive reference. All the rule must do is determine the reference of a token of the first-person pronoun, and it does this adequately, and without circularity.

Anscombe's reason for thinking the rule was inadequate was that it did not differentiate self-conscious reflexive reference from unintended reflexive reference. But whilst she is right to think that it does not, there is no semantic requirement that the rule do so. All that a semantic account of a term need do is account for the determination of reference of that term (Bermúdez, 1998, p. 17). To the extent that it does just that, the rule is sufficient. And since it is not insufficient, there is no need to supplement the rule with an indirect reflexive reading of the third person pronoun, and hence the account of first-person reference is not circular. The reference of the first-person pronoun is determined by the self-reference rule and the context of utterance: a token of the first-person pronoun refers to the producer of that token.

3.6 Self-consciousness and first-person thought

In Section 3.2 I suggested that we take first-person reference to be determined by the self-reference rule in conjunction with the context in which an utterance of the first-person

pronoun is made. The rule states that a token of 'I' refers to the individual who produces that token, whilst the context determines who the producer of the token is. This gives us an account of the contribution that a use of the first-person pronoun makes to the truth-conditions of an utterance of a sentence in which 'I' occurs. That is, the self-reference rule account of first-person reference is an account of the semantics of the first-person pronoun.

The account tells us something about what a subject needs to know if she is to understand utterances of 'I'. If a subject is to produce a comprehending utterance of 'I' to refer to herself then she needs to know the character or conventional significance of 'I', and we have given the self-reference rule as the character of 'I'. A subject must know that an utterance of 'I' will refer to the person who produced it. She might not characterize the character of 'I' in just this way, but she must know some analog of the rule if she it to produce comprehending utterances of 'I' in order to refer to herself. So too, someone who understands an utterance of 'I' that they hear must know that rule or its counterpart if they are to know who it is that is being referred to. But this is not all that a hearer of an utterance needs to know in order to understand that utterance, and neither is it all that a producer of an utterance needs to know if she is to produce comprehending utterances of 'I'.

With indexical terms quite generally, and with the first-person pronoun specifically, we have to posit some further knowledge on the part of speakers and hearers who use and understand these terms. Suppose you receive a postcard through the post on which is typed the message 'I am having a great time' (Perry, 1988; Millikan, 1990; DeVidi, 2001). As someone who is a competent speaker you understand the meaning of 'I'. You know that 'I' refers to the person who has produced it, and thus you understand 'I' as a *type* of referring term. You know, then, the character of the first-person pronoun: that 'I' contributes 'the agent of the context' to the truth-conditions of the sentence in which it occurs. But, as DeVidi points out, when you read the postcard, what you are interested in is knowing *who* it is that is having a great time. You don't just want to understand the sentence as a type of expression. Knowing the self-reference rule, or something like it, will be a necessary but not a sufficient condition on your understanding the token expression presented to you on the postcard (DeVidi, 2001, p. 42). And being able to think of the person who produced the written token of 'I' on the postcard as the person who produced the written token does not give you an understanding of that *token* or the particular token of the sentence in which it is used: it gives you only knowledge that it is the producer of the token that is the element of the context to which the utterance refers.

So, what would help you understand the token? The reader needs to have knowledge of

the context: they need to know who it was that typed the message. One needs to have some epistemic contact with the referent distinct from one's knowledge of him or her as simply 'the producer of the token of 'I''. We can contrast the example of your receiving a postcard from me through the post, with a situation in which I stand before you and say 'I am having a great time'. You can see me and hear my utterance. In this situation you know the context of utterance. You know that the person you can see is the person who has said 'I am having a great time'. What the visual experience of me gives you is epistemic access to the producer of the token of 'I' such that you can understand the utterance. The point is that you need to know what the context is: you need to be aware of the referent in an appropriate way so that you know that she is the agent of the context and therefore the reference of the token utterance. To interpret or understand an indexical, one must have some kind of knowledge of which thing stands in the appropriate relation to the utterance to be its producer (Millikan, 1990).

In a sense this is no less true of oneself when one speaks about oneself as it is of others when they understand one's utterances of 'I'. One does not, of course, typically have to work out that it was oneself who produced a token of 'I'. We do not interpret our own utterances of 'I' moving from context to referent, using the character of the first-person pronoun to determine ourselves as the referents of our tokens of 'I'. So, we do not usually need to know the context of the utterance to understand our own utterances, after the fact.

But, recall the lesson we learnt in examining Anscombe's objection to the self-reference rule account of first-person reference. When we produce an utterance of 'I' we know that we use the term as a term with reflexive reference, in other words, to speak about ourselves as ourselves. As O'Brien says, 'I'-use not only conforms to the self-reference rule as such. In addition we use the term *as* a term governed by the self-reference rule. This means that one must have some way of thinking about oneself such that one can produce a token of 'I' knowing that it refers to oneself.

A hearer uses her knowledge of the character of 'I' to take her from context to referent. So she must know the context as well as the character if she is to understand an utterance of 'I'. By contrast, a speaker does not use her knowledge of the character of 'I' to take her from context to referent. She uses her knowledge of the character of 'I' in conjunction with her intention to refer to herself to generate or create the appropriate context—that she is a producer of a token of 'I'. One intends to speak about oneself and, knowing that 'I' refers to the producer of an utterance of 'I', one produces an utterance of 'I'. What is required, then, to use the first-person pronoun, is a prior capacity for first-person thought.

O'Brien (1994, 2007) cites David Lewis's (1983a) example of two gods whose propositional knowledge is complete and who also know the self-reference rule, but who are not self-conscious to illustrate the dependence of the capacity for first-person reference on first-person thought:

"[…] one [god] lives on the tallest mountain and throws down manna and the other lives on the coldest mountain and throws down thunderbolts, both are omniscient with respect to propositional knowledge. Neither god knows which god he is. We can also take it that the gods know that 'I' refers to whoever uses it and that they know for any utterance of 'I' who made that utterance. So if the god on the tallest mountain utters 'I am throwing down manna' they both know that the 'I' refers to the god on the tallest mountain. Similarly if the god on the coldest mountain utters 'I am throwing down thunderbolts' they both know that the 'I' refers to the god on the coldest mountain. But it cannot be said that either god succeeds in referring to himself in the first person way. Thus it seems that the gods' understanding of 'I' as a device of self-reference is not sufficient for them to be able to use it to achieve self-conscious self-reference. Notice that the case hinges on taking the gods to be in some peculiar way estranged from their own utterances – they precisely seem not to know that they, themselves are uttering when they are. (Indeed this estrangement seems to me sufficient to rule out the possibility that the gods can be considered as genuine speakers). What the gods lack, but we do not, is self-conscious knowledge of our own utterances. If is that knowledge of our own actions which enables us to self-consciously self-refer when we use a term as governed by the self-reference rule." (O'Brien, 1994, pp. 280-281)

As O'Brien says, what is missing from the gods' uses of 'I', and what is required from us, is the "first personal identification of ourselves in thought" (O'Brien, 1994, p. 280). To use 'I' correctly and comprehendingly one must know that one is referring to oneself, and therefore one must have some knowledge of oneself. That is, to master the self-reference rule and thereby intentionally produce utterances of 'I' that refer to oneself, one must have some knowledge of which person one is, or some way of thinking about ourselves, which one employs in forming the intention to produce intentional utterances of 'I' (Bermúdez, 1998; Coliva, 2003). So, without an account of first-person thought, we cannot give an adequate account of our capacity for use of the first-person pronoun.

3.7 First-person reference and first-person thought

An obvious suggestion concerning first-person thought is that, just as the reference of 'I' is fixed by the self-reference rule, so the reference of the self-concept employed in first-person thought is fixed by a similar reflexive rule. The reference of a token of the self-concept is fixed by the rule that tokens of the self-concept refer to the thinker who produces the token. O'Brien (2007) considers this in her discussion of first-person reference. If first-person thought is to explain how it is that we have the capacity to use the first-person pronoun, then

we need to give an account of first-person thought in order to provide a full explanation of the capacity for first-person pronoun use. But, if our account of first-person thought—the employment of the self-concept in thought—is an account of the determination of reference in terms of a token-reflexive rule, then we will raise yet a further problem, because first-person thought is self-conscious thought (O'Brien, 2007, p. 69).

We saw in Chapter One that when one thinks a first-person thought one thinks about oneself self-consciously. Thus, we will need an account of how it is that we can think about ourselves by employing a concept whose reference is governed by a token-reflexive rule. We will need to posit some other, more basic, way of thinking about ourselves first-personally. Even if we take the reference of a thought that employs the self-concept to be fixed by a token-reflexive rule, we will need to give some account of how we know that we are thinking about ourselves when we employ that self-concept. What a subject seems to know is that she is entertaining a first-person thought, so we will need to ascribe to the subject a second first-person thought. As such, we still need to give an account of this other capacity to think about ourselves first-personally: we still need an account of first-person thought.

3.8 Conclusion

We saw in Chapter Two that we do not have observational or quasi-perceptual awareness of a self when we introspect. We also saw in that chapter that this gives us a problem in specifying the content of the self-concept, and thus with the determination of the reference of the first-person pronoun. In this chapter we have seen that, whilst we can understand first-person reference to be fixed by the self-reference rule, the account that we give will not be complete because it fails to provide an explanation of the self-consciousness of first-person thought. In Chapter Six I will return to the idea that reference is determined by a self-reference rule, in particular to the idea that the self-concept is individuated in terms of its reference rule, and examine how we might supplement the account so that it can accommodate the self-consciousness of first-person thought (Peacocke, 2010).

But first I will examine alternative accounts of how it is that first-person reference is determined, accounts designed to accommodate the fact that first-person thought is self-conscious thought because they offer accounts of the determination of reference in terms of the content of first-person thought. If introspection does not provide us with perceptual knowledge of a self or subject, then what kind of knowledge can we bring to bear in thinking about ourselves first-personally? One possibility is that descriptive knowledge based on our acquaintance with our own mental states supplies the content of our self-concepts. In the

next chapter I will consider whether descriptivism allows us to account for the content of the self-concept.

Chapter 4: Descriptivism and first-person thought

4.1 Introduction

In the previous chapter I looked at an account of the determination of first-person reference, according to which reference is determined by a reflexive rule that specifies the producer of a token utterance of the first-person pronoun as the referent of that token. The view that first-person reference is determined by the self-reference rule plus the context of production has several advantages: it accounts for the guaranteed referential success that we have in using 'I', and it respects the context-sensitivity of the first-person pronoun. 'I' is a term that refers to different individuals in different contexts, and the idea that reference is determined by the rule plus the context of utterance explains this context-sensitivity.

However, the view fails to account for the fact that our utterances of 'I' express self-consciousness. When we speak about ourselves using the first-person pronoun we do so with the knowledge that we are referring reflexively to ourselves. Moreover, the extension of the account to first-person thought only generates the same problem: we seem to need an account of a basic kind of first-person thought if we want to explain how we use a particular self-concept as a concept governed by the token-reflexive rule that a token of the concept refers to whoever produced it.

In this chapter I will consider an alternative that attempts to provide an account of how it is that we think of ourselves when we entertain first-person thoughts such that those thoughts are self-conscious, and moreover provides an account of how it is that first-person reference is determined. We established in Chapter Two that we do not observe or become acquainted with a self when we introspect, so it cannot be perceptual or quasi-perceptual information that characterizes the content of our 'I'-thoughts and thereby determines the reference of the first-person pronoun. But there is an alternative to perceptual knowledge—or its predecessor, knowledge by acquaintance—in the shape of knowledge by description.[18] Moreover, a common, though not uncontroversial, interpretation of Frege's semantics has it that sense is descriptive. Could it be, then, that descriptive knowledge characterizes the content of our self-concept?

[18] The distinction is controversial, but in this chapter I will assume that there is enough of an intuitive idea of two distinct kinds of knowledge that it is reasonable for us to consider the possibility that we think of ourselves first-personally 'by description'.

4.2 Descriptive sense

A common view of Frege's semantic theory takes the sense expressed by a linguistic term to be a description of the object that is the semantic value of the term.[19] On this view, the sense expressed by an utterance of 'I' would be a description of the person using the term, a description that is uniquely satisfied by the person using the term. Reference is mediated by sense: a term denotes a particular object because the term expresses a descriptive sense that specifies that object uniquely.

There are notable and well-established objections to such a view, stemming from Saul Kripke's *Naming and Necessity* (1980). If the sense expressed by a term is descriptive, then the truth-value of the proposition will be indexed to the possible world in which it is evaluated. The semantic value of the term will vary with relevant differences in the possible worlds in which the proposition is evaluated. In worlds in which the description denotes a different individual from the one denoted in the actual world, the semantic value will also be different from that in the actual world.

For Perry (1977) the problem with descriptive senses is that descriptive knowledge does not seem necessary for us to entertain first-person thoughts: we can entertain first-person thoughts even when we do not know any discriminating facts about ourselves. Moreover, it does not even seem sufficient: descriptive thoughts fail to have the cognitive significance and motivational force of our 'I'-thoughts.

First, I will look at the modal concerns with the idea that the first-person pronoun expresses a descriptive sense. The denotation of a definite description will depend on who or what uniquely satisfies the description in question, and this will differ in different circumstances, or possible worlds. If propositional content is descriptive and therefore the semantic value of a term that is used in the expression of that proposition is determined by that description, then the semantic value of the term will depend on the circumstances in which the proposition is evaluated. Propositions containing descriptions will have different truth-values—and the term that expresses that description will denote different individuals or objects—in relevantly different possible worlds.

Take the definite description 'the winner of the Wimbledon Men's Singles Final in

[19] Some of the motivation for interpreting Frege in this way seems to come from the following passage from the fourth footnote of 'On Sense and Reference': "In the case of an actual proper name such as 'Aristotle' opinions as to the sense may differ. It might, for instance, be taken to be the following: the pupil of Plato and teacher of Alexander the Great" (Frege, 1993, p. 24, n. 4). Evans (1982; 1985a) has argued for a contrasting interpretation of Frege according to which the sense of a referring expression is to be understood as a way of thinking of an object.

2011'. In the actual world the description denotes Novak Djokovic. But in a set of possible worlds Rafael Nadal won the singles final that year. In those possible worlds, the definite description 'the winner of the Wimbledon Men's Singles Final in 2011' denotes Nadal rather than Djokovic.

If the content of first-person thought is to be characterized by a description *D*, then it seems that a sentence using the first-person pronoun should express just the same proposition as a sentence in which description *D* is employed. The problem is that, in circumstances in which description *D* denotes a different individual to the one denoted by *D* in the actual world, then the propositions expressed by a sentence in which the first-person pronoun is used and a sentence in which *D* is employed will be assigned different truth-values.

Lets suppose that the descriptive content that we think an utterance of the first-person pronoun by Novak Djokovic expresses can be characterized by the description 'the winner of the Wimbledon Men's Singles Final in 2011'. So, we test whether sentence (1) below expresses a proposition with the same truth-value as the proposition expressed by sentence (2):

(1) I am Serbian.

(2) The winner of the Wimbledon Men's Singles Final in 2011 is Serbian.

When evaluated in the actual world both these sentences express propositions that are true, if uttered by Djokovic. Djokovic is Serbian, so his utterance of (1) expresses a true proposition. And since the description 'the winner of the Wimbledon Men's Singles Final in 2011' denotes Djokovic in the actual world, then (2) also expresses a true proposition. But what about the possible worlds in which Rafael Nadal is the unique individual denoted by the description 'the winner of the Wimbledon Men's Singles Final in 2011'? Nadal is not Serbian, but Spanish. So, in those possible worlds, Novak Djokovic's utterance of (1) will express a true proposition,[20] whilst his utterance of (2) will express a false proposition. And so, it cannot be the case that (1) and (2) express the same proposition. Hence, it seems that we must abandon the idea that we can characterize the content of the proposition or thought expressed by the first-person pronoun, in other words, the content of first-person thought, in terms of a definite description.

We get to the same conclusion if we embed the first-person in a modal context (Kaplan, 1989a, pp. 519-520). If 'I' expresses the descriptive content 'the winner of the Wimbledon Men's Singles Final in 2011' when 'I' is uttered by Novak Djokovic, then a sentence such as

[20] I assume here that being Serbian is an essential property of Novak Djokovic.

the following uttered by Djokovic would seem to express a true proposition:

(3) 'If no one was the winner of the Wimbledon Men's Singles Final in 2011, I would not exist.'

According to the descriptivist about first-person thought, and given the assumption that an utterance of 'I' expresses the descriptive content 'the winner of the Wimbledon Men's Singles Final in 2011', the proposition that (3) expresses can be characterized as follows:

(4) {If no one was the winner of the Wimbledon Men's Singles Final in 2011, the person who is the winner of the Wimbledon Men's Singles Final in 2011 would not exist}

Whilst (4) is trivially true, the proposition expressed by (3) is not true, for Novak Djokovic's existence is not dependent on there being a winner of a particular sporting tournament. The Wimbledon Men's Singles Tennis Championships might have been cancelled in 2011 without Djokovic failing to exist. The descriptivist must interpret sentence (3) as expressing proposition (4), but this cannot be right. So, again, we see that the content of first-person thought cannot be descriptive. Call these the modal objections to the descriptivist account of first-person thought.

There are a second set of reasons for rejecting a descriptive view of first-person thought. Perry (1977) objects to the view that indexical terms express descriptive senses on the grounds that having a set of beliefs about oneself is neither necessary nor sufficient for one to entertain 'I'-thoughts. As we already saw in Chapter One, a descriptive thought about oneself does not seem to have the cognitive and motivational significance that our first-person thoughts have. Moreover, one's beliefs about an object or individual seem to be irrelevant to the thought or proposition one expresses with an indexical term that refers to that object or individual.

First, Perry argues that one's beliefs about the referent of an indexical do not comprise part of the content of thought expressed by that indexical (Perry, 1977, p. 12). One might believe incorrectly, for example, that today is the 1st April 2012, whilst still entertaining thoughts that we naturally express using the indexical term 'today'. Suppose I say 'Today is April Fool's Day', believing that today is the 1st April 2012. But my belief is incorrect: it is in fact the 2nd April 2012. If the content of my thought were to comprise the description '1st April 2012', then my utterance would express a true proposition. But this cannot be right: in uttering the sentence 'Today is April Fool's Day' on the 2nd April 2012 I have expressed a false proposition. My beliefs, whether correct or incorrect, seem to be irrelevant to the thoughts that I express when using indexical terms.

Moreover, having any beliefs at all seems to be unnecessary for one to have the capacity

to entertain first-person thoughts (Perry, 1977, p. 12). The point can be illustrated by way of Anscombe's (1994) sensory-deprivation tank thought experiment (Howell, 2006, p. 45). Suppose that I am suspended in a sensory deprivation tank, with no inner or outer bodily awareness and no current experience of the world. Suppose also that I am suffering from amnesia so that I have no knowledge of my personal history, my name, and so on. Even if I am in such a position, I am able to make the following kind of utterances: 'I hope I never get myself into this situation again', 'I wonder where I am', and so on. And in making those utterances, I express truth-evaluable thoughts or propositions, even though I have no set of beliefs about myself. Perry takes this to show that the content of thought expressed by the first-person pronoun cannot be descriptive. The non-necessity of descriptive knowledge for one to be able to entertain 'I'-thoughts is indicated by the fact that one can entertain 'I'-thoughts without having any descriptive knowledge of oneself. Call this the argument from non-necessity of belief.

Furthermore, Perry argues that merely holding a set of true beliefs about the individual one happens to be will not suffice to enable one to entertain first-person thoughts (1977, pp. 12-13). The descriptive thoughts one can entertain about oneself will not be first-personal, lacking the cognitive significance and motivational force of first-person thoughts, as I described in Chapter One. Perry's example of Rudolf Lingens in the Stanford Library can be used to indicate this insufficiency (Howell, 2006, p. 45):

> "An amnesiac, Rudolf Lingens, is lost in the Stanford library. He reads a number of things in the library, including a biography of himself, and a detailed account of the library in which he is lost. He believes any Fregean thought you think might help him. He still will not know who he is and where he is, and no matter how much knowledge he piles up, until that moment when he is ready to say,
>
> *This* place is aisle five, floor six, of Main Library, Stanford.
> *I* am Rudolf Lingens." (Perry, 1977, p. 17)

Rudolf Lingens might have the belief, based on his reading, that the man in aisle five on floor six of Main Library, Stanford at a particular time is Rudolf Lingens. If the first-person pronoun is used to express a descriptive sense, then the description 'the man in aisle five on floor six of Main Library, Stanford' would be a good candidate for the descriptive content of the first-person thought expressed by an utterance of a sentence in which 'I' occurs. But, even when Lingens has the description available to him, and knows that the individual that satisfies that description bears the name 'Rudolf Lingens', he will not act appropriately if he is told that Rudolf Lingens must report to the reception desk at the library. If the description provided Lingens with a first-person way of thinking of himself he would respond to hearing the message by walking to the reception desk of the library. But of course, Lingens would

not respond in this way to the message, because the thought that the man in aisle five on floor six of Main Library, Stanford is Rudolf Lingens does not provide Lingens with a way of thinking about himself self-consciously. That this thought is distinct from the thought that Lingens would express with the sentence 'I am Rudolf Lingens' is also evidenced by the fact that Lingens can rationally adopt different epistemic attitudes towards the descriptive thought and the first-person thought (Evans, 1985a, p. 292). Lingens believes the descriptive thought, but does not believe the first-person thought. Descriptive thoughts about oneself lack the cognitive significance and motivational force of first-person thoughts. Hence, descriptions do not capture the content of first-person thought. Call this the argument from insufficiency.

Perry's arguments are compelling: he is surely right to claim that we are not provided with a descriptive specification of ourselves of the kinds considered above every time we use the first-person pronoun, and that even where we do have discriminating factual knowledge of ourselves, those descriptive facts do not provide the content of the thought that is expressed by a use of the first-person pronoun. Having descriptive knowledge of ourselves of the kind that Perry considers does not provide us with the means to entertain 'I'-thoughts.

In this chapter I am considering whether we think of ourselves descriptively when we entertain first-person thoughts. Perry's arguments suggest that this cannot be right: the content of one's self-concept cannot be descriptive. The worry is that the contents of one's 'I'-thoughts are not specified by any description or set of descriptions, no matter how thoroughgoing. But, there is a further kind of descriptive thought that seems to be a better candidate to characterize the content of first-person thought, and it is to this that I will now turn.

4.3 A description of the subject of thought

According to Robert J. Howell (2006), in spite of the worries raised by Kaplan and Perry, we should give descriptivism another chance.[21] Howell claims that we come to the conclusion that descriptions cannot characterize the contents of our 'I'-thoughts because we fail to take into consideration a particular description that is much better placed to capture the

[21] Howell is concerned with defending the certainty of the *cogito* from Hume's thesis about the elusiveness of the self to introspective awareness. To do this he aims to give what he calls a "[…] psycho-semantics, or a semantics of thought, which analyses the concepts involved in basic self-knowledge so that the notable epistemic features of the cogito are explained by the mechanisms through which we most fundamentally refer to ourselves" (Howell, 2006, p. 46).

motivational force of our 'I'-thoughts than the ones so far considered. By taking the description we use to think about ourselves to involve reference to our own mental states, Howell aims to answer Perry's concerns over the adequacy of descriptions in capturing the content of our 'I'-thoughts. Howell also argues that we can take the descriptive content of first-person thought to determine the reference of an utterance of 'I'. He suggests that we can avoid the modal objection by stipulating that there is a difference between the public content expressed by an utterance of 'I'—which is a singular proposition—and the private, descriptive content entertained by the subject when she uses 'I'. The general proposal is that the kind of description we should consider is to be articulated in terms of one's acquaintance with one's mental states. Our knowledge of our mental states is sufficiently secure to make it a plausible basis for a descriptive self-concept.

Howell's starting point is Descartes' *cogito*. When Descartes questions his own existence, mistrusting the deliverances of his senses, he finds one thing of which he cannot doubt the existence: his own doubt. Descartes' meditator, after admitting that he cannot trust his senses and therefore that he has no certainty about the existence of the world and even of his own body, questions whether he could also be deceived about his very existence. It is possible that an evil genius has tricked him into believing in the existence of his body and the world. Could it be the case that an evil genius is controlling his thoughts such that he has been persuaded he exists, even though he does not?

"But there is a deceiver of supreme power and cunning who is deliberately and constantly deceiving me. In that case I too undoubtedly exist, if he is deceiving me; and let him deceive me as much as he can, he will never bring it about that I am nothing so long as I think that I am something. So after considering everything very thoroughly, I must finally conclude that this proposition, *I am*, *I exist*, is necessarily true whenever it is put forward by me or conceived in my mind." (Descartes, 1986, p. 17)

When the meditator considers whether he could be tricked about his own existence, it is his own doubt that shows him that he does exist. Descartes proclaims that as long as he thinks, then he does indeed exist. Even if Descartes doubts his existence, the having of the doubting thought is sufficient to show him that his whole existence is not an illusion. He has evidence of his own existence as long as he thinks:

"Thinking? At last I have discovered it – thought; this alone is inseparable from me. I am, I exist – that is certain. But for how long? For as long as I am thinking. […] But for all that I am a thing which is real and which truly exists. But what kind of a thing? As I have just said – a thinking thing." (Descartes, 1986, p. 18)

The only piece of knowledge that Descartes can trust is his knowledge of his thought, his doubt. He has rejected his sensory experience of his body and the world as providing

evidence of existence that can withstand radical scepticism. Therefore he cannot think of himself as the individual that satisfies a description that makes reference to his body or the world. The only thing he feels any certainty about is his thoughts. Yet he is able to think of himself:

"It is by his mental states that he knows himself, and as such it should be to them we look for descriptions by which we can self-refer." (Howell, 2006, p. 49)

So too, an amnesiac in a sensory deprivation tank knows nothing about his body. Recall Anscombe's thought experiment concerning the ability of an individual suspended in a sensory deprivation tank to entertain thoughts employing the self-concept. We have already considered adding to the thought experiment the stipulation that the individual is suffering from amnesia (Howell, 2006, p. 45; see also Evans, 1982, p. 215). He has no observational knowledge of his body, and no memories of his past. As such, it is not possible for him to think of himself under a description formulated in terms of bodily states or past events, and this leads us to the conclusion that the thinker has no descriptive knowledge of himself and of course, no observational knowledge either. But this is not in fact the case. As with Descartes' meditator who is unable to trust his senses, the amnesiac undergoing sensory deprivation has introspective knowledge of his own mental states (Howell, 2006, p. 49). Howell claims that this first-person knowledge of his mental states provides the basis for descriptive knowledge of the subject of those mental states. It is, therefore, on the basis of one's awareness of one's mental states that Howell frames the descriptive content of the self-concept. He considers two possibilities, which he labels type descriptivism and token descriptivism.

According to type descriptivism, the content of one's self-concept is a description of oneself as the thing that is thinking whatever type of thought one is currently thinking. Take Descartes' performance of the *cogito*. Descartes thinks the thought expressed by the sentence 'I am thinking'. According to type descriptivism, Descartes' self-concept has the content we might represent with the phrase 'the thing that is thinking I am thinking'.

"I refer to myself in that very statement as that which is thinking the proposition <I am thinking>." (Howell, 2006, p. 49)

The description specifies the content of my 'I'-thoughts by giving a description of the person who is the thinker of a particular type of mental state, the mental state that I currently enjoy. Even if I could no longer remember my past, feel my body or sense the world, so long as I could still think at all, I would still have acquaintance with my thoughts, and so would still have the capacity for descriptive knowledge of myself as the thinker of my thoughts.

The self-concept I might employ in thought might be characterized by the description 'The thing that thinks I am no longer able to remember my past, feel my body, or sense the world', if my current thought about myself is expressed by the sentence 'I am no longer able to remember my past, feel my body, or sense the world'.

But type descriptivism runs into problems just because the properties specified in the descriptions are types, meaning that they can be had by anyone. The description 'the thing that is thinking I am thinking' is not a definite description. It cannot be guaranteed to individuate a unique individual (Howell, 2006, p. 49). Some of our thoughts might be entirely original and specific to ourselves, but many of our thoughts will not be. Take the *cogito*. As Howell points out, it is possible that there are several philosophers performing the *cogito* at the same time. The description 'the person thinking I am thinking' is not a proper definite description, and thus does not hook on to a unique referent.

To avoid this problem Howell proposes what he calls token descriptivism instead. I think of myself not as the thing enjoying a type of mental state, as type descriptivism suggests, but as the thing enjoying a particular mental state. The content of one's self-concept will then be characterized by the description 'the thing with *this* thought' or 'the thinker of *this* thought'.[22]

The term 'this' in that content is supposed to refer demonstratively to the mental state with which one is acquainted. Hence, the thought that I am happy might have the content:

{The thinker of *this* thought; the property of being happy}

Token descriptivism avoids the failure of guaranteed unique reference that is a problem for type descriptivism since, although others may entertain the very same type of thought as I currently do, no one else can be said to entertain my thoughts. As such, the description picks out a unique referent. It is not just the type of mental state or mental property that is used in describing the self, but the mental state or property itself that is referred to demonstratively within the description.

Particular mental states will be suitable to be included in the descriptive knowledge that can be used to determine the referent of the self-concept if they are the objects of conscious attention. When I attend to my mental state introspectively, that mental state can be 'imported into the description'. Howell takes this position to follow Bertrand Russell's in 'On the Nature of Acquaintance' (1914a-1914c):

[22] Howell's formulation of the description by which we can think of ourselves self-referentially is reminiscent of one of P.F. Strawson's forms of identifying reference: "For even though the particular in question cannot itself be demonstratively identified, it may be identified by a description which relates it uniquely to another particular which can be demonstratively identified" (Strawson, 1959, p. 21).

"When something is known by acquaintance, the thing is directly known without intermediaries, and descriptions can then be built using those things as part of their constituents. Conscious mental states or experiences are paradigm objects of acquaintance, and they are imported into the content of a proposition by the mental act of attention associated with the demonstrative <this>. I can refer to myself as that which is having this sensation where the object of attention itself is a part of the proposition." (Howell, 2006, p. 50)

4.4 Sense, reference and the modal objection

Howell's account offers a way of answering the question we started with: what is it for a subject to entertain first-person thoughts? His proposal is that a description of the subject as the subject of a particular token of a mental state characterizes the content of first-person thought. No subject is presented when we introspect, but since our mental states are available to awareness, we refer to ourselves as the thinker of those thoughts, referred to demonstratively.[23] Note that the token descriptivist account respects the indexicality of the first-person pronoun: reference will differ as the context of utterance differs. My utterance of 'I' will refer to the thinker of the thought expressed by the sentence in which 'I' is used, a thinker who is picked out by demonstrative reference to the thought that I entertain. Each descriptive content will include a demonstrative reference to the particular thought of which it is a part. And hence, reference will shift with the context.

Howell's theory of the content of first-person thought is a theory about the sense expressed by utterances of the first-person pronoun. As such, it is a theory about how it is that first-person reference is determined. The descriptive content of first-person thought expressed by an utterance of 'I' determines the reference of an utterance of 'I'. Specifically, the description 'the thinker of *this* thought' is employed by a subject when she thinks about herself and is expressed by utterances of the first-person pronoun such that the term can be

[23] There are two possible interpretations of Howell's account. The account on offer has obvious similarities with the mixed descriptive-demonstrative characterization of the content of first-person thought proposed by Christopher Peacocke in *Sense and Content* (1983). But, Howell differentiates his view from Peacocke's, saying that his own view is Russellian, whilst Peacocke's is Fregean. Presumably, then, Howell does not take the content of first-person thought to be characterized by the content:

{The thinker of *this* thought, the sense of the predicate phrase '…is happy'}

It would seem that Howell instead advocates the view that an utterance of a sentence like 'I am happy', expresses three general propositions:

{Something is thinking this thought}
{Only that thing is thinking this thought}
{The thing is happy}

However, nowhere does Howell explicitly characterize the content of first-person thought in this way. I will therefore focus my attention on the Fregean interpretation of the descriptivist account.

said to pick out the individual that the description denotes. The term does not refer directly, but rather is mediated by the description:

"[…] self-reference is underpinned by descriptive knowledge about the self, enabled by one's consciousness of one's own sensations." (Howell, 2006, p. 46)

When I produce an utterance of 'I', then the semantic value of my utterance will be me, because I uniquely satisfy the description 'the thinker of *this* thought'. But, by offering a descriptive account of the content of first-person thought as that which determines the reference of an utterance of the first-person pronoun, Howell's account appears to be vulnerable to the modal objection which I described in Section 4.2, a fact which Howell himself recognizes. For example, the propositions expressed by the sentences

(5) I am writing a PhD thesis

(6) The thinker of *this* thought is writing a PhD thesis

will be assigned the same truth values in the actual world, but at a possible world at which I am writing a PhD thesis, but I am not the thinker of the demonstrated thought, then they will be assigned different truth-values (Howell, 2006, p. 57). It seems that the thought expressed by my utterance of (5) cannot be a descriptive thought with the content 'the thinker of *this* thought'.

Moreover, the proposition expressed by the sentence

(7) If no one were to think *this* thought, then I would not exist

will be true if the first-person pronoun contributes the description 'the thinker of *this* thought' to the content of the proposition expressed by the sentence in which it is used, because that proposition will be as follows:

{If no one were to think this thought, then the person who thinks this thought would not exist}

But, of course, a person would not fail to exist if one of their thoughts did not exist (Howell, 2006, p. 56). Again, it seems that we cannot take the content of first-person thought to be descriptive.

Howell's response is to adapt his account to include a directly referential aspect to it. The usual response to the modal objection is to take the first-person pronoun to contribute the individual that is the referent to the content of the proposition expressed by an utterance of a sentence in which 'I' is used. First-person thought, on this view, is singular thought, and the first-person pronoun is directly referential. So, for Kaplan (1989a), my utterance 'I am happy' expresses the propositional content:

{Alisa, the property of being happy}

Howell wants to adopt this view of propositional content, so as to avoid the problems arising for the descriptivist when indexical terms are embedded in modal contexts. He does not, however, want to relinquish his descriptivist account altogether. Direct reference theories have problems of their own: they have difficulty explaining cognitive significance.

Suppose that I am suffering from amnesia and no longer know my own name. The identity statement 'I am Alisa' seems to be informative. According to Kaplan proper names and indexicals contribute the objects or individuals to which they refer to the proposition expressed by utterances of sentences containing those terms. So, the identity statement, for Kaplan, expresses the trivial proposition:

{Alisa, the property of being Alisa}

If we take the propositions expressed by sentences containing indexicals to be singular, then we have a problem explaining why it is that the identity statement is informative. Direct reference accounts of indexicals need to be supplemented in some way if they are to be explanatorily adequate. As we saw in Chapter Three Kaplan suggests that we can account for cognitive significance in terms of character, the conventional meaning of the term that determines who the referent will be in a context. Suppose that you and I both believe that I am about to be attacked by a bear. I believe that I am about to be attacked by a bear, and you believe that I am about to be attacked by a bear. What we both believe is the proposition:

{Alisa, the property of being attacked by a bear}

But, we believe it under different characters. I believe it under the character of the first-person pronoun, whilst you believe that same proposition under the character of the third-person pronoun (Perry, 1977, pp. 18-19). It is not clear exactly what believing or entertaining a proposition under a character is supposed to involve, but even aside from that worry, there is a problem with the approach. Character does not seem to be fine-grained enough to explain cognitive significance (Howell, 2006, p. 59). On the direct reference view the demonstrative term 'this' will contribute the object to which it refers to the propositional content expressed by the sentence in which it is used, and cognitive significance is to be explained by the character of the term. Suppose that we are looking at the Royal Yacht Britannia in the harbor at Leith.[24] The stern of the yacht is visible to one side of a large building, whilst the bow—with the name 'Royal Yacht Britannia' painted on it— is visible to the other. The middle of the ship is hidden from view. When you point to the bow of the ship and say '*This* is the Royal Yacht Britannia' I accept what you say, but when you say '*This* is the Royal Yacht Britannia' pointing to the stern, I do not believe you. According to

[24] The example is adapted from John Perry's (1977).

the direct reference theorist, the same proposition is being expressed by each sentence. What is more, that proposition is being entertained under the same character: the character of 'this'. Yet I am rational even if I hold different epistemic attitudes to the proposition expressed by your first utterance and the proposition expressed by your second utterance. It does not seem that character can do the work needed to explain cognitive significance.

Howell wants to use his token descriptivist account to explain cognitive significance, and the direct reference account to avoid the modal objection. According to two-tier descriptivism an utterance of the first-person pronoun will express both a singular proposition, which he calls the public content, and a private, descriptive content characterized as 'the thinker of *this* thought'. The worry is that directly referential terms must denote the same object in all possible worlds, and that terms whose semantic value is determined by a descriptive sense will not meet this condition since the semantic value will be indexed to the world at which the proposition is evaluated, and will vary depending on how things are at each world. By specifying that the 'public content' of one's thought is a singular proposition, Howell aims to avoid the modal objection: in all possible worlds my first-person thought will refer to me, Alisa. But, I will also entertain a private content, with the descriptive content 'the thinker of *this* thought', and it is this that explains why an utterance of the sentence 'I am Alisa' will be informative if I am suffering from amnesia.[25]

However, we will see in Section 4.6 that Howell's token descriptivist private content does not provide an explanation of the cognitive significance of first-person thought after all. The description 'the thinker of *this* thought' does not seem to capture the cognitive significance of 'I'-thoughts any better than other kinds of descriptions that might be used to specify the content of first-person thought.

But first in Section 4.5 we will see that his proposal leads to a proliferation of first-person contents: not only do we postulate the existence of public and private contents, but each distinct first-person thought will have its own unique first-person content.

4.5 The problem of the proliferation of self-concepts

Howell's account gives the content of first-person thought as a description in which demonstrative reference is made to one's current mental state. Each first-person thought,

[25] We can also make the same point by claiming that the first-person pronoun will express a primary and a secondary intension: the primary intension being the descriptive content 'the thinker of *this* thought' whilst the secondary intension will be the individual who is the referent of the token of the term.

then, will have a different content, and we will have difficulty explaining how it is that we are able to think about ourselves over time and make valid inferences about ourselves over time. Each person will think about herself at a particular time in such a way that she will never think about herself in the same way again.

Howell avoids problems that arise for type descriptivism—the problem of the description failing to pick out a unique individual—by specifying that the description that characterizes the content of one's 'I'-thoughts involves demonstrative reference to particular token mental states. But this raises a different problem for the account: different token mental states will generate different descriptions, which suggests that the content of first-person thought will be different for each thought one entertains about oneself. Even if one's mental state now and one's mental state five minutes ago are of the same type, one's demonstrative reference to each state will mean that the description that fills out the first-person content of the thought expressed by an utterance of 'I', will be different for each state. As a result, the number of distinct first-person thoughts will multiply rapidly. For these self-concepts to play any role in our reasoning about ourselves as substantial, enduring beings we will need to supplement the content of the self-concept.

Howell acknowledges something similar to this problem, which arises from the 'thin, momentary grasp' we have of 'ourselves' when we are acquainted with a single mental state:

> "[…] since we identify ourselves by mental tokens which do not endure for long, we lose any ground for reidentifying ourselves, thus losing any grounds for asserting our personal endurance through time. […] My answer […] is rather straightforward: the conditions under which we self-refer in basic cogito-like instances are not the only cases of self reference, nor is knowledge of ourselves as thinking beings the only case of self-knowledge. There are undoubtedly other cases of self-knowledge that do not aspire to the cogito level of certainty (and that do not survive radical Cartesian doubt) but that still constitute knowledge, and it is these cases that would provide security in reidentification. This would seem to be exactly the result we desire. Cogito-type knowledge doesn't seem to provide us with reidentification criteria for ourselves. Such criteria surely comes posterior to basic Cartesian reflection." (Howell, 2006, p. 54)

Howell states his aim as being limited to giving an account of self-reference when we have no other knowledge of ourselves other than our acquaintance with our current mental state. Of course such limited knowledge, he argues, will not provide us with a more robust and lasting concept of ourselves.

But our designs are grander: we want to know how it is that we think of ourselves such that we can reason about what to do and think about how we have acted. The consequence of there being multiple distinct self-concepts is that we cannot account for the validity of the inferences we make that employ those distinct self-concepts in different premises (Bar-On,

2004, pp. 34-5). For example, we typically think that an inference such as the following is valid:

I was thinking about the *cogito*
I am thinking about descriptivism

I was thinking about the *cogito* and am thinking about descriptivism

But if there are in fact two distinct first-person contents used in premise one and premise two, then how can we account for my inference from those premises to the conclusion that I was thinking about the *cogito* and am thinking about descriptivism? Howell's account suggests the following propositional content should be assigned to the first two sentences above:

{The person thinking *this very thought*, the property of thinking about the *cogito*}
{The person thinking *this very thought*, the property of thinking about descriptivism}

Given our aims, it is clear that a description such as 'the thinker of *this* thought' is not sufficient to characterize the content of our self-concept. What we need to add to this description is a conception of an enduring subject. The minimal self-conception we get from the description will not allow us to think about ourselves as anything more than an individual that exists for the length of time that the mental state exists. With each new mental state that we become acquainted with we are able to generate a new self-concept. But this does not give us an account of anything substantial and enduring. In order to reason about ourselves using these slim self-conceptions we will need to supplement those self-conceptions with a conception of ourselves as something that endures through time.

Of course, if the description 'the thinker of *this* thought' does not give us a satisfactory self-conception, and we must look elsewhere for a conception of ourselves as enduring, then we return to our initial question: how do we conceive of ourselves? So far we have seen that we do not think about ourselves as subjects of thought on the basis of observation or acquaintance. And in this chapter we have seen that even if we can think of ourselves momentarily by means of the description 'the thinker of *this* thought', these descriptions will not be sufficient to characterize a conception of the thinker as a reasoning, enduring subject.

4.6 The cognitive significance of 'I'-thoughts

The need to supplement Howell's description with something more also becomes apparent

when we think about whether his account is successful in meeting Perry's arguments from non-necessity and insufficiency. Howell argues that a subject who is able to entertain first-person thoughts will always have available to her the information required to entertain a thought with the content 'the thinker of *this* thought', so his account can meet the argument from non-necessity. What is more, he claims that his version of the descriptivist account of first-person thought has the appropriate cognitive significance to meet Perry's concerns about insufficiency.

However, although we will see that Howell's account is not susceptible to the argument from non-necessity, it will become clear that the account must be supplemented if the descriptive thoughts that Howell proposes are to have the cognitive significance of first-person thoughts. What we shall see is that the descriptive content 'the thinker of *this* thought' will "prove either redundant or inadequate" (O'Brien, 2007, p. 69). The problem is that a thought with the descriptive content that Howell suggests will fail to have the cognitive significance and motivational force of a first-person thought unless the subject is able to identify the demonstrated thought as her own.

As we saw at the start of this chapter Perry argues that holding true beliefs about oneself is neither necessary nor sufficient for one to entertain 'I'-thoughts. One's knowledge of a set of descriptive facts about oneself, or lack thereof, seems to be irrelevant to one's capacity to think 'I'-thoughts. So too, having a thorough set of beliefs about the person who I happen to be will not enable me to think about myself first-personally. There is a gap between Rudolf Lingens's belief about himself that he expresses with the description 'the person in aisle five, floor six, of Main Library, Stanford' and his belief about himself that he expresses using the first-person pronoun. Descriptive knowledge does not seem to be required, nor does it provide sufficient conditions to allow us to entertain 'I'-thoughts.

Howell's challenge is that the description 'the thinker of *this* thought', where 'this' refers demonstratively to a mental state with which the thinker is acquainted—presumably the very mental state in which the description occurs—is not only available to a subject in all situations in which she is able to entertain first-person thoughts, but also carries the cognitive significance of our 'I'-thoughts:

> "Token descriptivism provides, I think, palatable and explanatory solutions to the semantical puzzles as well. With respect to the amnesiac in the sensory deprivation tank, we have located information that is present whenever the cogito is possible. Performance of the cogito may not require much, but it does require that one be conscious and have occurrent mental activity, and this is all that token descriptivism requires for self-reference. Furthermore, the description appears to be of the right sort to solve the second puzzle and Perry's exposition of it. When Rudolph Lingens is lost in the Stanford library, it seems that once he adds to his expansive knowledge the descriptive information formulated in terms of

his current sensations, he has sufficiently located himself." (Howell, 2006, pp. 50-51)

As a response to the argument from non-necessity, Howell's descriptivist account of first-person thought is promising. However, as it stands, the account fails to provide a response to Perry's argument from insufficiency. The phenomenon of thought insertion experienced by patients suffering from schizophrenic delusions seems to suggest that the description 'the thinker of *this* thought' does not have the full cognitive significance that 'I'-thoughts have, and that possession of descriptive knowledge of the sort described by Howell is not sufficient to enable one to entertain an 'I'-thought (O'Brien, 2007, pp. 69-70).

First, I will look briefly at Howell's answer to the argument from non-necessity. Perry argues that one need not have descriptive knowledge in order to entertain 'I'-thoughts, which becomes apparent when we realize that we can retain the capacity to entertain first-person thoughts in the absence of descriptive knowledge. For example, with the thought experiment of the amnesiac in the sensory-deprivation tank case we seem to have a situation in which descriptive knowledge is lacking and yet the subject can still entertain 'I'-thoughts. Hence, descriptive knowledge is not necessary for one to entertain 'I'-thoughts, and it cannot be the case that the self-concept is descriptive. But, as we have already seen in Section 4.3, Howell argues that in the scenario described the subject still has *some* descriptive knowledge. Even in cases where much of our descriptive knowledge is lacking we still have a thin or minimal descriptive knowledge of ourselves. As long as we continue to be acquainted with our mental states by being able to entertain them—which is presumably necessary for us to entertain 'I'-thoughts at all—we seem to be able to think of ourselves as 'the thinker of *this* thought'. So, it seems not to be the case that we can entertain 'I'-thoughts in the absence of all descriptive knowledge. Our capacity to think 'I'-thoughts seems to coincide with our ability to think of ourselves by way of the description 'the thinker of *this* thought'. Howell's descriptivism undermines Perry's argument from non-necessity. The argument rests on the intuition that we can entertain first-person thoughts when we have no descriptive knowledge of ourselves. But Howell points out that even in the kind of imaginary scenarios that we use to test out intuitions about first-person thought, we still have descriptive knowledge of ourselves based on our acquaintance with our own mental states.

What about Howell's claim that descriptive knowledge is sufficient for one to be able to entertain a first-person thought? Howell also thinks the description 'the thinker of *this* thought' can be used to meet Perry's objection that descriptive knowledge is not sufficient for us to be able to entertain 'I'-thoughts.

Perry argues for the insufficiency of descriptive knowledge on the grounds that thinking

a descriptive thought about oneself is not a way of thinking of oneself first-personally. Rudolf Lingens's thought that the person in aisle five, floor six, Main Library, Stanford is Rudolf Lingens is not a first-person thought in which he identifies himself as the bearer of the name 'Rudolf Lingens'. His descriptive thought lacks the cognitive significance and motivational force of first-person thoughts. The problem is that when one entertains a first-person thought one knows one is thinking about oneself: one's thought is self-conscious. But thinking a thought with the descriptive content 'the person in aisle five, floor six, Main Library, Stanford' is not a way of thinking of oneself such that one knows that one is thinking about oneself.

Howell's account can be seen as a response to Perry's argument from insufficiency: unlike other descriptions that do not provide Lingens with a first-personal way of thinking of himself, Howell takes the description 'the thinker of *this* thought' to imbue the thoughts in which it occurs with the cognitive significance and motivational force of first-person thoughts. The thought that the thinker of *this* thought is Rudolf Lingens is, says Howell, a self-identifying thought (Howell, 2006, p. 51). If Lingens is told that Rudolf Lingens should report to the reception desk of the Main Library and he knows that the thinker of *this* thought is Rudolf Lingens, where 'this' refers demonstratively to the very thought that Lingens is entertaining, then Howell seems to think that Lingens will act appropriately, walking to the reception desk of the Main Library.

Howell's suggestion amounts to the claim that 'I' is synonymous with the definite description 'the thinker of *this* thought'. As such, an individual that holds differing epistemic attitudes to the proposition expressed by a sentence in which 'I' is used and the proposition expressed by a sentence in which the description 'the thinker of *this* thought' is used, where the property ascribed is the same in each proposition, should be considered irrational. It should not be possible for a subject to ask a question such as the following: the thinker of *this* thought is thinking about the problem of the self, but am *I* thinking about the problem of the self?

Even though the description 'the thinker of *this* thought' seems more suitable as a characterization of the content of first-person thought than a description such as 'the person in aisle five, floor six, of Main Library, Stanford', upon consideration we see that thoughts with such a content will still lack the cognitive significance that Howell claims them to have. O'Brien objects to this kind of approach on the grounds that a description in which one makes demonstrative reference to one's mental state will turn out to be either redundant or inadequate (2007, p. 69). The description will be inadequate to capture the cognitive

significance and motivational force of our 'I'-thoughts, and the supplementation that is needed to render the account satisfactory will also render the description—and in particular the demonstrative reference to one's mental state—redundant. A descriptive thought will only have the cognitive significance and motivational force of first-person thought if that description itself employs the first-person. But by incorporating the first-person into the description we render the demonstrative reference to one's mental state superfluous, and also make the account circular. We should, then, abandon the idea that the content of first-person thought can be characterized by the description 'the thinker of *this* thought'.

The problem with Howell's characterization of the content of first-person thought in terms of the description 'the thinker of *this* thought' is that it seems possible that one might rationally accept the proposition expressed by a sentence using that description, but deny or remain agnostic about the proposition expressed by a sentence that employs the first-person pronoun, or vice versa. In particular, this possibility seems to arise for those who experience the phenomenon of thought insertion. Some schizophrenic subjects report that they entertain thoughts that are not their own, or for which they are not the agents. The sufferer is aware of, or acquainted with, those mental states, yet she denies something like authorship or production of those states.

But this suggests that a person suffering from thought insertion can rationally doubt the proposition expressed by the sentence 'I am thinking *F*', say, whilst accepting the proposition expressed by the sentence 'The thinker of *this* thought is thinking *F*', and whilst retaining the capacity for first-person thought (Peacocke, 2010, p. 90). The experience seems to be exactly the kind of experience that supports the subject in asking, 'the thinker of *this* thought is thinking F, but am I thinking F?'

The schizophrenic is delusional in taking her thoughts to be the thoughts of some other person who has inserted them into her awareness or mind. Yet, the phenomenon does suggest that the description 'the thinker of *this* thought' does not characterize the content of first-person thought, or does not do so without supplementation. After all, how could the schizophrenic be entertaining first-person thoughts that are distinct from the descriptive thought with the content 'the thinker of *this* thought' if the latter content is supposed to characterize the content of the former?

For those suffering from thought insertion what seems to be missing is the identification of themselves with the thinker of their current mental state. They are acquainted with the inserted mental states, but they do not think that they themselves are the author or thinker of the thought. We might, then, take Howell's descriptive content 'the thinker of *this* thought'

to be a part of the content of first-person thought, augmented by an identification of the thinker of the demonstrated thought with something else. In which case we can explain the phenomenon of thought insertion in terms of the identity component of one's self-individuation coming unstuck from the descriptive content 'the thinker of *this* thought'. It is for this reason that schizophrenic subjects disown their own thoughts, and also for this reason that it is rational to hold differing epistemic attitudes to descriptive thoughts and first-person thoughts. The simple descriptive thought with the content 'the thinker of *this* thought' is not a first-person thought, since it lacks the identity judgement that would give the thought the cognitive significance of a first-person thought.

Thought insertion indicates the inadequacy of demonstrative reference to one's mental states to generate genuine 'I'-thoughts. But, we have not yet seen that the descriptive content 'the thinker of *this* thought' is not a *part* of the content of first-person thought. What seems to be missing—and required for a description to characterize an 'I'-thought properly—is the identification of the mental state as one's own. So, the description must be something like 'the thinker of *this* thought, where *this* thought is *my* thought' (O'Brien, 2007, pp. 69-71).

But if this is right, then we might wonder what work the description is doing in characterizing the self-concept, as O'Brien (2007, pp. 69-71) points out. For what seems to be needed is the capacity to identify oneself with the posited thinker of the mental state with which one is acquainted, which means that we must posit a prior capacity to think about oneself. One's acquaintance with one's mental states enables one to think about the thinker of that thought, but one needs to identify that thinker with oneself if the thought is to have the cognitive significance of a first-person thought. And that identity judgement will itself employ a first-person content, so our account appears to be circular.

Now, it might be the case that one's acquaintance with one's mental states will be sufficient to provide one with the knowledge that the thought one is acquainted with is one's own thought. If it is sufficient, then it seems that demonstrative reference to the thought will in fact be unnecessary, and something that we can eliminate from the description to give us the content 'the thinker of my thought' (O'Brien, 2007, p. 69). But again, this is hardly a satisfactory characterization of the content of first-person thought, containing within it the concept for which we are trying to provide an account. What is the point of characterizing the content of the self-concept in terms of a description like 'the thinker of *my* thought' when such a description contains reference to oneself? One's being able to entertain a thought with this kind of descriptive thought will rest on one having a prior capacity to think about oneself.

Moreover, the existence of individuals who experience thought insertion suggests, anyway, that it cannot be acquaintance with one's mental state that provides grounds for the identification of the state as one's own. The fact that such individuals do not identify as their own the thoughts that they experience as inserted suggests that mere acquaintance does not provide one with the grounds to think of one's thoughts as one's own.

The description 'the thinker of *this* thought' appears to be inadequate to characterize the content of first-person thought, as the phenomenon of thought insertion suggests. Yet, an addition to that descriptive content that makes it adequate both renders the demonstrative reference to one's current mental state superfluous, and also renders the account of first-person thought circular.

4.7 Conclusion

In this chapter I have been considering whether the content of the self-concept can be identified with a description. Howell suggests that the content of thought entertained by a subject who utters the first-person pronoun can be characterized by the description 'the thinker of *this* thought', where 'this' refers to a particular mental state with which the subject is acquainted. We have seen, however, that there are objections to this suggestion. On the one hand, the account cannot provide us with a satisfactory account of how we think of ourselves over time, since there will be a multitude of contents—potentially a different concept for each mental state with which one is acquainted. As such, the account is not satisfactory without supplementation.

Furthermore, Perry's argument from insufficiency does not seem to be surmountable without supplementation that renders the account circular. We need to supplement the description with an identification of the mental states with which one is acquainted as one's own. But, as O'Brien points out, even if one's acquaintance with one's mental states provides one with the grounds to identify those mental states as one's own, the descriptive content under consideration would be something like 'the thinker of *my* thought'. Such a content would ensure that one's thoughts have the motivational force of one's first-person thoughts, but this would come at the expense of the account being circular for we would have an unexplained use of a first-person content within the description used to characterize the content of first-person thought. Furthermore, the phenomenon of thought insertion suggests that acquaintance with one's mental states alone does not ground the identification of those thoughts as one's own.

Chapter 5: An objective self-concept

5.1 Introduction

I started out by considering first-person thoughts in which we ascribe psychological states, and noticed that we seem to make such self-ascriptions on the basis of introspection alone, at least much of the time. But when I considered whether we observe a subject to which our mental states belong, I discovered both that, as it happens we do not observe a self, and that, on consideration, it seems that observation of a subject of thought is not possible. We seem to make psychological self-ascriptions on the basis of introspection alone, but there is no introspective observation of a subject or self.

However, there is an obvious way in which we do observe ourselves. We might not observe any psychological subject of experience when we become acquainted with our mental states, but we are in receipt of various kinds of perceptual information about the material body. The self-concept might therefore be a concept of a material object—the body—and we might satisfy the possession conditions for that self-concept by being party to perceptual information about the material body. According to Gareth Evans (1982) we can give an account of first-person thought by combining a functional characterization of the self-concept with a set of constraints that a subject must meet if she is to entertain a thought at all (1982, p. 212, pp. 262-263). Relevant to our interests, in particular, is Evans's insistence that a subject must have discriminating knowledge of the object about which she is thinking if she really is to entertain a singular thought. In the case of first-person thought that object is oneself, and to discriminate oneself from all other objects of one's kind one needs knowledge of one's temporal and spatial location in the objective world. Evans argues that all first-person thought must comprise either practical knowledge of one's objective spatial location or, at least, knowledge that one is a material being with temporal and spatial location if it is to count as genuine thought. For Evans one's self-concept is a concept of an object—a bodily being who has a spatial location in the world.

Evans claims that we do not entertain first-person thoughts in which we ascribe psychological properties to ourselves on the basis of introspection only. Of course, we do entertain thoughts in which we predicate psychological properties to ourselves. However, for Evans it is a condition on first-person thought that we think of ourselves always as material objects, and we satisfy this condition by having the capacity to locate ourselves spatially and

temporally on a cognitive map, either through being in receipt of perceptual information that enables such self-location, or at least by being disposed to use such information if it is forthcoming.

There are several objections that I will make to the objective approach to the self-concept. The first objection arises from the tension between perceptual demonstrative accounts of first-person thought and the intuition that the first-person has guaranteed reference. It seems that even when perceptual information is attenuated, a subject will be able to think about herself first-personally. A view that explains the self-concept in terms of perceptual information must either claim that our perceptual faculties are unfailing, or else must reject the intuition that we might mistakenly think about someone else or no one at all when we employ that concept.

Evans seeks to avoid the problem of guaranteed referential success by modifying his account: it is not that one's self-conscious thought about oneself is all and only thought informed by particular kinds of perceptual information about oneself: rather, one is *disposed* to treat those kinds of perceptual information as germane to one's first-person thought. The problem, though, is that this response leads us to the conclusion that we have not yet provided an account of the self-concept, as O'Brien (1995, 2007) argues.

Secondly, the approach requires too much sophistication on the part of the subject to be plausible. Evans requires that his subjects be able to locate themselves on a cognitive map built up through engaging in conceptual spatial reasoning based on perceptual information from their environments. But we can imagine subjects who do not have the capacity to perform such a high-level cognitive task, yet are able to entertain first-person thoughts. Indeed, it seems possible that young children might develop the capacity to entertain first-person thoughts or proto-first-person thoughts before they have the cognitive competence to engage in conceptual spatial reasoning.

In response to this objection I will describe an alternative: José Luis Bermúdez (1998) offers an account of proto-first-person thought in terms of a subject's possessing a non-conceptual point of view on her environment, in part explained by the subject's having developed an integrated representation or cognitive map of her environment and her own place in that environment. It will be a necessary though not a sufficient condition on a subject possessing such an integrated cognitive map of her environment that she be able to navigate around her environment in particular ways. Bermúdez's account has similarities to Evans's account, in particular in offering an account of first-person thought in terms of a capacity for self-location and an awareness of oneself as a material object. Bermúdez offers

an objective account of first-person thought, but his account differs from Evans's because he does not require that a thinker can engage in the conceptual spatial reasoning that Evans demands of his subject.

Finally, I will consider what seems to be the strongest objection to the objective approach to the self-concept. This final objection gets at the heart of the idea that we must think of ourselves as material beings if we are to think of ourselves first-personally. The conceptual coherence of the thought that one is in fact disembodied shows that it cannot be that we always think of ourselves as material objects when we entertain first-person thoughts. We need a concept of ourselves as thinker and to revise our idea of the discriminating knowledge required for successful first-person thought so that we can allow that a subject can think of herself without thinking of herself as a material object.

5.2 Evans's perceptual demonstrative account

In Chapter Two I rejected a perceptual demonstrative account of first-person thought on the grounds that we do not introspectively observe a self or subject of thought when we come to know about our mental states and experiences. In this chapter I will consider an alternative, third-person account of the self-concept. Our capacity to think about ourselves is based on ways of knowing or becoming aware of ourselves as spatio-temporally located individuals. When we think about ourselves we think of ourselves as objects. True, there is no inner object for us to perceive and refer to with a quasi-demonstrative first-person concept, but there is an external object that we do perceive in the shape of our material bodies (O'Brien, 1995, p. 233). It is through our knowledge of ourselves as material beings that we have the capacity to entertain first-person thoughts about ourselves.

I will take as the paradigm for this approach Evans's account of self-identification in *The Varieties of Reference* (1982). Evans's account cannot be classed as a straightforward perceptual demonstrative account of first-person thought, as we shall see, for there are two components to the account he gives. On the one hand, Evans individuates the first-person concept in terms of the relations that thoughts in which it is employed stand in to particular ways we have of gaining information about ourselves, and the relations those thoughts stand in to appropriate action; on the other hand, Evans's account of first-person thought must accord with his general framework for successful singular thought—a subject will only entertain a first-person thought if she knows which object it is about which she is in receipt of information (Evans, 1982, p. 212). The two components can neither be reduced one to the

other, nor can they be entirely separated from one other (1982, pp. 262-3). By standing in an appropriate informational relation to the object one is thinking about one can meet the knowing-which requirement Evans places on singular thought. Being in receipt of particular kinds of information about herself will give a subject the capacity to locate herself in space and time, thereby meeting the knowing-which requirement on successful thought. I will come to an examination of this component in the next section. First, I will describe the first component of Evans's account: the functional characterization of the first-person-concept.

The account of first-person thought that Evans offers corresponds to the structure of the account he gives of demonstrative thought. In giving his account of demonstrative thought Evans aim is to explain how it is that demonstrative thought depends on perception, or how it is that perception enables us to entertain thoughts about particular objects in the world (1982, pp. 143-145). To this end, he introduces the notion of an information channel. As subjects of experience we stand in informational relations to the objects we experience: we are in receipt of information about those objects. The information one receives about an object one perceives provides one with a conception of that object. This conception controls or governs one's thinking about that object.[26] By this Evans means that a subject will be disposed to treat the information she receives about the object as immediately relevant to the truth-value of the thoughts she entertains about that object. If I think a demonstrative thought about the book that I see, then the conception of the book that governs my thinking will be determined by the content of that perception.

Evans's account of the first-person concept is, in part, an account of the role that first-person thought plays in the cognitive dynamics of a subject. First-person thought is thought that is related to particular ways we have of gaining information about ourselves, but it is also thought that is manifested in appropriate behaviour. Evans recognizes that both aspects are required to give a thorough functional characterization of first-person thought, but he focuses on the connection between first-person thought and our ways of finding out about ourselves (1982, pp. 206-7). Like demonstrative thoughts, first-person thoughts are thoughts that are controlled by information that we receive by standing in information-links with

[26] Evans uses the term Idea for the component of thought that I have been calling a concept. I will continue to use the term concept in this chapter for the sake of continuity with the rest of the thesis. It is important, though, to realize that my use of the term concept cannot be interchanged with Evans's term 'conception'. A governing conception is not to be equated with the Idea or concept employed in thought. A conception of an object comprises the deliverances of the information channel that connect subject to object. As we will see, such information will not always provide the subject with an adequate concept. To equate all demonstrative Ideas of objects with the conceptions the subject has of those objects is to fail to recognise the importance of meeting the knowing-which requirement on thought.

ourselves. We have a variety of ways of finding out about ourselves, only some of which provide input into the conception of ourselves that controls our first-person thinking.

Evans uses identification-freedom as the test for an information channel's contributing to one's controlling conception of oneself, and he gives the following criterion as a test for a thought's being identification-free:

"What we should say is that a judgement is identification-free if it is based upon a way of knowing about objects such that it does not make sense for the subject to utter 'Something is *F*, but is it *a* that is *F*?', when the first component expresses knowledge which the subject does not think he has, or may have, gained in any other way." (Evans, 1982, pp. 189-190)

The criterion is familiar from Shoemaker's (1968) 'Self-Reference and Self-Awareness', where it is used as a test for immunity to error through misidentification relative to the first-person pronoun. Evans uses it as a test for identification-freedom, where he takes a successful thought in which an adequate concept is employed and which is identification-free to be immune to error through misidentification.[27] For Evans, the test for identification-freedom is something like a test for sameness of sense (Grush, 2007). It is not rational for a subject to hold different epistemic attitudes to what are identical thoughts. So, for example, if *a* and *b* are senses that determine the same referent, *O*, and I believe that *a* is F, but doubt that *b* is F, I will only be rational if *a* and *b* are distinct ways of thinking about *O*. If *a* and *b* are in fact the very same mode of presentation of object *O*, then it is irrational or nonsensical for me to question whether *b* is F, but assert that *a* is F, or vice versa. Similarly, if it does not make sense or it would be irrational for a subject to ask 'Something is F but is it *c* that is F?', where the first phrase expresses knowledge gained through the deliverances of a single information channel, then the information from that channel provides the controlling conception of the object thought about, when one employs the concept *c*.

But we cannot individuate concepts or modes of presentation of an object in terms of different information channels. Several information channels can all contribute to one's controlling conception of an object. The test for identification-freedom, then, can be used as an indicator that a disposition on the part of the subject to entertain thoughts employing a particular concept-type on the basis of the information received from an information channel is part of a functional characterization of that concept-type (Evans, 1982, p. 183). The question 'someone is F, but is it me that is F?' is to be used to discover which of the ways we have of gaining knowledge about ourselves are ways of gaining information about ourselves upon which our first-person thoughts depend.

[27] I will explain the conditions Evans places on successful thoughts and adequate concepts in the next section.

One receives proprioceptive information about the position of the parts of one's body relative to the rest of the body. Say, for example, that one has a proprioceptive experience that provides one with information about one's legs being crossed. Now, says Evans, it does not make sense to ask 'Someone's legs are crossed, but is it my legs that are crossed?' when one entertains the first proposition—'someone's legs are crossed'—on the basis of proprioceptive information. As such, one's thought that one's legs are crossed will be identification-free. What this shows is that proprioceptive information is one of the ways we have of gaining knowledge of ourselves on which first-person thought depends. First-person thoughts are thoughts that are controlled by proprioceptive information about the body.

But proprioception does not exhaust the ways we have of gaining information about ourselves which dispose us to entertain first-person thoughts. We have other ways of gaining information about our bodies 'from the inside':

"[…], we have what might be described as a general capacity to perceive our own bodies, although this can be broken down into several distinguishable capacities: our proprioceptive sense, our sense of balance, of heat and cold, and of pressure." (Evans, 1982, p. 220)

In addition to these information channels, we also receive information about our bodies through perception of the world, and our self-ascriptive thoughts made on the basis of such information are also identification-free. It makes no sense to ask 'Someone is standing in front of the library, but is it me who is standing in front of the library?' when the first proposition ('someone is standing in front of the library') is entertained on the basis of one's current visual perception of the library only. So perception of the world constitutes an information channel, or a set of information channels, between a subject and herself such that she is disposed to treat information received via that channel, or those channels, as immediately relevant to the truth or falsity of her first-person thoughts.

We also gain information about ourselves as psychological beings, although it should be noted that Evans rejects a strict distinction between mind and body that the separation of the two kinds of knowledge here might suggest. We come to know about our epistemic attitudes and experiences, and both types of self-ascriptions—when they are made on the basis of deciding what to believe and the conceptualization of our experience in thought—will be identification-free. Similarly, our memory-based self-ascriptions will be identification-free.[28] It does not make sense to ask 'Someone believes it is raining, but is it me that believes that it is raining?', and neither does it make sense to ask 'Someone saw a tree burning, but is it me

[28] I have here given only a very brief description of Evans's account of mental and memory-based self-ascription, which does not do justice to the significance or subtlety of what Evans has to say on such matters. The importance and details of his account lie beyond the concerns of my thesis.

that saw a tree burning? All these ways of gaining information about ourselves contribute to the conception of ourselves which controls or governs our first-person thinking.

5.3 An objective self-concept

The structure of the account Evans offers of first-person thought also follows the structure of demonstrative thought in having a second component: a subject must know which object it is that she knows about through the information channels listed in the previous section (Evans, 1982, p. 212). According to Evans this knowing-which condition will be met if a subject has the capacity to locate herself in space and time. A subject needs this capacity if she is to satisfy two principles that Evans sets out as governing successful thought about the world.

The first condition he labels Russell's Principle: one can only refer to an object in thought if one has discriminating knowledge of it, where one has such knowledge if one can distinguish the object from all other things of its kind. One must know which particular thing or individual in the world one is thinking about (Evans, 1982, p. 89). In order that one's thought latches onto an object, it is necessary to individuate the object in thought, which requires that one employs a discriminating conception of it, a conception that enables one to distinguish that object from others of its kind. Russell's Principle is often called a 'knowing-which' requirement on thought. By satisfying this condition on thought, a thinker can be said to know what it is for the thought she entertains to be true, and thus can be credited with the thought. Russell's Principle applies to all thought, but, of course, this means that our attempts at entertaining first-person thoughts must also conform to Russell's Principle if they are to count as successful thoughts.

The second condition or principle Evans introduces he calls the Generality Constraint. The Generality Constraint stipulates that if one is to understand a particular thought predicating property F of object *a*, then one must be able to entertain thoughts that predicate other properties of *a*, and also be able to entertain thoughts that predicate *F* to other objects (Evans, 1982, p. 104, p. 209).

> "[…] any thought which we can interpret as having the content *that a is F* involves the exercise of an ability—knowledge of what it is for something to be *F*—which can be exercised in indefinitely many distinct thoughts, and would be exercised in, for instance, the thought that *b* is *F*. Similarly for the thought that *a* is *G*." (Evans, 1982, p. 103)

What is required if one is to understand the thought that *a* is F, is that one should understand what it would be for objects other than *a* to be F, and for object *a* to instantiate properties other than F.

We can satisfy both Russell's Principle and the Generality Constraint if we employ an adequate concept or Idea of an object in thought.[29] Russell's Principle requires that we have a discriminating conception of the object about which we are thinking—that we know how to distinguish it from all other things of its kind. Most straightforwardly we have such a discriminating conception of the object we are thinking about when our thought is at the 'fundamental level'. At the fundamental level of thought we employ fundamental identifications of the objects we think about, which involves having knowledge of the fundamental ground of difference of that object—that which distinguishes the object from all other things of its kind.[30] According to Evans, the fundamental ground of difference for material objects is their spatio-temporal location. To think a fundamental thought about a particular material object one must have the capacity to locate that object in space and time. A fundamental thought will employ a fundamental concept or Idea.

Not all out thoughts are at the fundamental level: we can think of objects even when we do not identify them fundamentally as the possessors of the fundamental ground of difference they in fact possess—as objects of a particular kind located at such-and-such a place and such-and-such a time. But entertaining non-fundamental thoughts will depend on having a certain kind of knowledge. We can only successfully entertain thoughts at the non-fundamental level if we know what it would be for the object we are thinking about to be identified fundamentally. If *a* is a non-fundamental identification of an object, I can only entertain the thought that *a* is *F* if I know what it would be for the propositions $\delta = a$ and δ is *F* to be true, where δ is an arbitrary fundamental Idea or concept of an object (Evans, 1982, pp. 108-9). This means that I can only employ concept *a* in thought to think about a particular object if I know generally what type of fundamental ground of difference that object has. Since the fundamental ground of difference for a material object is its spatiotemporal location, to think about oneself non-fundamentally, one must know that the object thought about would be identified fundamentally by knowing its location in space and time.

So, an adequate concept of a material object will enable one actually to locate the object in space and time, or alternatively will comprise knowledge that the way in which one would make a fundamental identification of the object about which one is in receipt of information

[29] I will not go into the details of why the Generality Constraint demands that we employ an objective concept when we entertain thoughts about material objects as I think the arguments take us beyond the requirements of this chapter.

[30] Evans employs the term fundamental identification to talk of the individuation of an object in thought, and not to refer to an identity judgement that would render the thought identification-dependent.

is by having knowledge of its spatial and temporal location.

Now, we saw already that Evans offers a broadly demonstrative account of first-person thought, according to which first-person thought is to be understood in terms of the role that such thought plays in the cognitive dynamics of that individual. First-person thought is understood to sit at the intersection of particular ways we have of gaining information about ourselves, and actions that are appropriate given the information we have about ourselves and the world. The self-concept is individuated by its relation to particular kinds of evidence we might receive about ourselves, and its relation to particular sorts of behaviour we might engage in. As with a pure demonstrative thought, then, we might think that it is through the information that we receive about ourselves that we can meet the knowing-which requirement and thereby also satisfy the Generality Constraint. People are material beings, so the fundamental ground of difference for a person will be her spatio-temporal location. Do the particular types of information we gain about ourselves described in Section 5.2 provide us with the capacity to locate ourselves in space and time?

Certain kinds of information which contribute to the controlling conception that a subject has of herself do seem to provide the subject with the capacity to locate herself in space and time. When one perceives the world, one not only receives information about the location of objects relative to oneself, one is in receipt of information about one's own location relative to the objects that one perceives. By engaging in spatial reasoning on the basis of one's perceptions of the world one will be able to locate oneself in public space. Evans gives the following account of the spatial reasoning one must be able to carry out on the basis of perceptual experience in order to locate oneself in space:

"'I perceive such-and-such, such-and-such holds at *p*; so (probably) I am at *p*'; 'I perceive such-and-such, I am at *p*, so such-and-such holds at *p*'; 'I am at *p*, such-and-such does not hold at *p*, so I can't really be perceiving such-and-such, even though it appears to me that I am'; 'I was at *p* a moment ago, so I can only have got as far as *p*', so I should expect to perceive such-and-such'. These arguments exploit principles connecting the subject's position, the course of his perceptions, and the speed and continuity of his movement through space; and the child must learn to trip round and round those principles, so that he comes to think effortlessly in these ways." (Evans, 1982, p. 223)

By having a theory of perception and action and the capacity to navigate around the world a subject will be able to construct a cognitive map of the world in which she not only locates the objects that she perceives, but also herself relative to those objects.

"The very idea of a perceivable, objective spatial world brings with it the idea of the subject as being *in* the world, with the course of his perceptions due to his changing position in the world and to the more or less stable way the world is. The idea that there is an objective world and the idea that the subject is *somewhere* cannot be separated, and where he is is given by what he can perceive." (Evans, 1982, p. 222)

A subject, then, will build up a cognitive map of the world using the perceptual information she receives and her capacities to reason about her own location relative to the objects she perceives. She will then be able to update this map and keep track of her own location using the continued deliverances of those perceptual information channels.

Proprioceptive and kinaesthetic information about the subject's own body might also provide the means for a subject to locate herself. A subject who has constructed a cognitive map of her environment will have the capacity to locate herself when she moves about even if certain other perceptual information channels are disrupted. If I know that I am standing facing the door of the room I am in before being blindfolded I will know where I am when I move around that room using kinaesthetic information to relocate myself on the cognitive map that I have of my environment. When I take three steps forward I will know that I am three steps closer to the door, even though I cannot update my location in space relative to the door using visual information at that time. Of course, in ordinary situations in which our perceptual faculties are all functioning normally kinaesthetic and proprioceptive information will also have a role to play in enabling us to keep track of our location in space. Ordinarily we keep track of where we are by drawing on the deliverances of various perceptual information channels at once.

It is not clear that some of the other information channels that contribute to one's controlling conception of oneself provide one with the means for one to make a fundamental identification of oneself. Information about the temperature of one's body, say, does not seem to provide one with the resources to locate oneself on a cognitive map, and neither does the information one receives about one's mental states. But it is possible that Evans envisages that the first-person concept will comprise a fundamental identification of oneself only on some occasions, and on other occasions will comprise a non-fundamental identification of oneself, depending on the evidence upon which a particular first-person thought is based.

We saw that at the fundamental level of thought the information that makes up one's controlling conception of an object gives one the capacity to locate oneself in space and time. But we also saw that an adequate concept need not comprise a fundamental identification of an object of this sort. An adequate concept will comprise, if not a fundamental identification, at least knowledge that the object of which one has a controlling conception is a material object and, therefore, that, were the right kinds of inputs forthcoming, the object would be identified by its location in space and time. Even when we are not in receipt of appropriate information to locate ourselves we retain our capacity for

spatial reasoning and our knowledge that such reasoning enables us to locate ourselves in public space. We employ this capacity and knowledge to think of ourselves non-fundamentally when we are not privy to the kind of perceptual information that would enable us to make use of our capacity for self-location. The first-person concept, then, is objective: it is a concept of a person located in the objective world. This will be the case when the information controlling one's first-person thinking provides one with the means to locate oneself in space and time, and also when the information fails to endow one with that capacity. In the latter case, in order to ensure that the first-person concept will be adequate and thereby meets the knowing-which requirement, one must at least have knowledge that one is a material being which has some location in space and time. As such, one's self-concept will always be a concept of a material being.

This account of first-person thought has several outcomes. As a neo-Fregean, Evans endorses the view not only that the sense expressed by a term determines its reference, but also that sense can be equated with the content of thought. His account of the first-person concept is thus an account of the reference of the first-person pronoun.

What is more, Evans thinks that it is a mistake to think that we can self-ascribe psychological properties on the basis of introspection alone. Some of our self-ascriptive judgements are based on introspection. However, Evans aims to show that it is a mistake to think that we employ a special and mysterious purely-introspection-based self-concept when we make introspection-based self-ascriptions. The capacity to entertain such introspection-based judgements is dependent on another capacity—the capacity to think about oneself as a spatio-temporally-located person. It is only by being able to identify myself as an element of the objective order that I am able to successfully entertain a thought. Whether I am ascribing psychological or material properties to myself, I must employ an objective self-concept in my thought if I am to succeed in entertaining a thought at all.

5.4 The potential failure of our perceptual faculties

Evans's account is a version of a general approach to first-person reference according to which perception provides the epistemic basis for reference (O'Brien, 1995, p. 232). Russell's Principle imposes an epistemic requirement on the subject. A term will refer if the subject who produces the term stands in an appropriate epistemic relation to the referent. This epistemic requirement is met by the subject's being in receipt of perceptual information about herself. But how might a subject meet the knowing-which requirement if her

perceptual systems are impaired?

If a simple version of the perceptual demonstrative account is correct, then, says O'Brien (1995, 2007) it seems that the absence of the relevant information should render the subject unable to entertain first-person thoughts. The simple version of the view would treat the self-concept as relevantly similar to demonstratives such as the 'this'-concept and the 'that'-concept. For such demonstrative concepts, the content of the concept is supplied by the current deliverances of a particular perceptual information channel. As long as we treat the self-concept as a demonstrative concept of this type, then we will have problems explaining how it is that we use the first-person concept when we have no information about ourselves. Moreover, if the subject, in some strange set of circumstances, were to receive information about someone else's body rather than her own, then it seems that she should employ the self-concept to entertain thoughts about someone other than herself. But, on the contrary, first-person thought seems to be extremely robust and, in particular, seems to be achievable even if one's perception of oneself were to break down or one's perceptual systems were to malfunction. Uses of the self-concept appear to be immune both to reference failure and to mistaken reference. When one uses the self-concept the referent of the concept seems to be guaranteed both to exist and to be appropriately present. Moreover, one seems to be guaranteed to refer to the right thing—oneself (Bar-On, 2004, p. 72). And this seems to hold true even when one misperceives, or if one does not perceive oneself at all. Our capacity for first-person thought seems to be independent from the successful functioning of our perceptual systems (O'Brien, 2007, p. 33).

Our bodily information systems might fail to provide us with information about ourselves but still provide us with information about another individual in just the way we would expect to receive information about ourselves. As David Armstrong suggests:

"We can conceive of being directly hooked-up, say by transmission of waves in some medium, to the body of another. In such a case we might become aware e.g. of the movements of another's limbs, in much the same sort of way that we become aware of the motion of our own limbs." (Armstrong, 1984, p. 113)

Thus, we can conceive of becoming aware of the world from the perspective or point of view of another person. Through some sort of transmission device I visually perceive what another individual sees, rather than the world around my body. As such the patterns of spatial reasoning I engage in which usually allow me to located myself in space will inform me about the location of the person about whom I am receiving information, and not about

my location.[31] If Evans's account is the simple perceptual demonstrative account, then he should say that my 'I'-thoughts would concern the source of information that I receive, rather than being about me (O'Brien, 2007, p. 39). But this does not seem to be the case. Even were I to receive perceptual information about the world from another person's perspective, it does not seem that my utterances of 'I' will refer to the individual that is the source of that perceptual information. Rather, it seems that I will still refer to myself when I employ the self-concept.

Anscombe's (1994) sensory deprivation tank thought experiment provides an example of another way in which we might fail to receive information about ourselves through the usual perceptual channels. In her thought experiment Anscombe asks us to imagine a situation in which we have no bodily or sensory information. She thinks that even when in the sensory deprivation tank we will still be able to refer using the first-person. As such, the types of information that are absent when in the tank cannot underlie our ability for first-person thought. It cannot be the case that we come to be able to entertain first-person thoughts by being presented with our bodies—there is no presentation of the body whilst in the sensory deprivation tank. Moreover, there is no perception of the environment, and hence no current input enabling us to locate ourselves on a cognitive map constructed from information about the location of objects in our environment. It seems that Evans, were he to advocate the simple perceptual demonstrative account, should be forced to accept that a subject in such a scenario would be unable to entertain first-person thoughts. But, again, this does not seem to be correct. Even if we were to receive no perceptual information about the environment and our bodies at all, then it seems we would still be able to think first-person thoughts (O'Brien, 1995, p. 235).

The possibility that we receive misinformation or no information about our environments, bodies and so forth, suggests that the simple perceptual demonstrative account cannot be correct. What can a proponent of the view say in response to the claim that our perception of the world can fail whilst our capacity for first-person thought remains?

5.5 Dispositions and circularity

Evans's way of dealing with this objection is to liken the self-concept to the demonstrative

[31] Evans seems to consider the possibility but does not say much in response. He suggests that a subject made aware of such a situation should moderate his thinking accordingly, but that we should not take the possibility of such a situation as informing our account of ordinary first-person thought (1982, pp. 249-250).

'here'-concept. Both first-person thoughts and 'here'-thoughts can be entertained without the subject currently being in receipt of any information about the referent of the concept (Evans, 1982, pp. 215-216). One need be in receipt of no current information about the place at which one is at for one to have the capacity to entertain 'here'-thoughts. So too, one need have no current information about oneself, and yet have the capacity to entertain first-person thoughts. Evans does not class 'I'-thoughts with demonstrative thoughts such as 'this'-thoughts as exemplary information-based thoughts, for which one's thinking employs a concept that is governed by the information currently received through an informational-link. Rather a subject will have the capacity to entertain 'I'-thoughts as long as she is disposed to have her first-person thinking controlled by information which might emerge from any of the relevant channels described above. A subject can employ the self-concept even if there is no information available to her through the specified channels at the time at which she thinks the first-person thought.[32]

Now, as we have seen, Evans sets up his account as having two interconnected aspects: the functional characterization explains the content of the self-concept in terms of the kinds of information that the subject might receive about herself along particular information channels, whilst the knowing-which requirement imposes certain constraints on the way a subject must able to think about herself if her thought is to be successful. To entertain a thought, a subject must know which person she is by knowing where she is located in space and time. We have seen, though, that these two aspects are interrelated: the information a subject receives about herself provides the subject with the knowledge of which person she is, by giving her information about her spatio-temporal location (Evans, 1982, p. 263).

By giving a dispositional reading to this account of first-person thought Evans creates a complication. An information channel can provide one with knowledge of which person one is as long as one is in receipt of suitable information through that channel. An information channel that does not provide one with any information does not give one knowledge of which individual one is. The subject in the sensory deprivation tank has no current perceptual information on which to base her knowledge of her location in her environment. Yet, she is disposed to have her thinking controlled by such perceptual information. How could it be that this disposition will be sufficient for a subject to entertain successful

[32] The same holds for the action-link for first-person thought—just as the information-link might fail to provide one with information about oneself, one can lose the capacity to act appropriately on the basis of one's first-person thoughts without losing the capacity to entertain first-person thoughts. A person who is paralyzed can entertain and understand the first-person thought expressed by the sentence 'I am about to be attacked by a bear', even though she will not perform any of the set of actions we would expect someone to perform were they to entertain and understand that thought.

thoughts about herself, thoughts in which she makes a fundamental identification of herself, or at least knows what it would be to identify herself in that way? Perhaps she can remember being in receipt of perceptual information along those channels, and she uses her memory to provide her with the knowledge of what it would mean to locate herself in space and time (Evans, 1982, p. 254). After all, something of this sort must be going on when we make a non-fundamental identification of ourselves, which I have suggested seems to be the case for many of the thoughts in which the self-concept is employed. A subject must know that she is the kind of thing that is located in space and time if she is to think genuine thoughts about herself when she lacks the practical capacity for self-location. But we can adapt the thought experiment, as Evans himself does, to add a further impairment: the individual in the sensory deprivation tank is suffering from amnesia and cannot remember having perceptual information about her environment such that she could locate herself with respect to the objects that she sees (Evans, 1982, p. 215). For the subject's first-person thoughts to be genuine thoughts she must think of herself as possessing the fundamental ground of difference that she possesses, or know what it would be to identify herself in such a way. So, if a subject is to be able to think about herself first-personally just as long as she is disposed to have her thinking controlled by particular perceptual information were it to be present, without actually receiving any information along such channels and without any memory of having received information along those channels, then we need some way of ensuring that the subject meets Russell's Principle.

O'Brien suggests that Evans can only get around this problem by claiming that a subject will know that there is one object to which she is dispositionally related in this way (O'Brien, 1995, p. 237). But this concession then leaves us with a further question: what basis does the subject have for her knowledge that she is dispositionally related in this way to just one object (O'Brien, 1995, p. 238)? In the absence of all perceptual information about the world and the body, in the absence of information from memory also, the subject is able to entertain first-person thoughts because she knows that there is just one particular object, information from which she is disposed to treat as immediately relevant to her first-person thinking. The problem we have is in accounting for the knowledge that she is related to just one object in this way, since all the information which could provide the subject with such knowledge is absent:

"In particular it is not clear how we are going to answer this question without assuming that the subject, independently of the actual receipt of information, has a capacity for self-reference which is involved in their knowledge that *they* are disposed to accept certain kinds of information as relevant to *them*." (O'Brien, 1995, p. 239)

It seems that we must introduce a capacity for first-person thought and a self-concept that is independent of the Evansian first-person thought and objective self-concept in order to explain how it is that we can entertain the latter type of thought and employ the objective self-concept. And given Evans's rejection of the idea that we can entertain first-person thoughts and employ a concept on the basis of nothing more than introspective awareness, we will have a problem explaining this mysterious primitive first-person thought that allows us to know that we are informationally related to just one object such that we are disposed to treat information that we might receive along particular information channels about that object as immediately relevant to our first-person thinking.

5.6 Sophisticated thinkers and Bermúdez on the non-conceptual point of view

We might object to Evans's account for a second reason also. The account is implausibly strong, making thought the achievement of only those who can satisfactorily engage in conceptual spatial reasoning.

The nature of the spatial reasoning that Evans demands of his subject seems to require too much of the subject. Evans requires that in order to think about oneself that one be able to conceive of oneself as an object among objects, with a spatial location. He gives the following account of the spatial reasoning one must be able to carry out on the basis of perceptual experience in order to locate oneself in space:

> "'I perceive such-and-such, such-and-such holds at *p*; so (probably) I am at *p*'; 'I perceive such-and-such, I am at *p*, so such-and-such holds at *p*'; 'I am at *p*, such-and-such does not hold at *p*, so I can't really be perceiving such-and-such, even though it appears to me that I am'; 'I was at *p* a moment ago, so I can only have got as far as *p*', so I should expect to perceive such-and-such'. These arguments exploit principles connecting the subject's position, the course of his perceptions, and the speed and continuity of his movement through space; and the child must learn to trip round and round those principles, so that he comes to think effortlessly in these ways. […] In 6.3 I argued that [our knowledge of what it is for us to be located at a position in space] in turn can be regarded as consisting in a practical capacity to locate ourselves in space by means of exactly the kinds of patterns of reasoning that I have just described. It is this capacity which enables us to makes sense of the idea that we ourselves are elements of the objective order; and this is what is required for our thoughts about ourselves to conform to the Generality Constraint." (Evans, 1982, p. 223)

Evans seems to suggest that those who have not yet acquired an understanding of these principles cannot have a self-concept, and therefore will be unable to entertain first-person thoughts. First-person thoughts will only be accessible to those who have a relatively sophisticated understanding of the relation between their perceptions and the world. One must be able to think of oneself as an element of the objective order. If first-person thoughts

are only available to those who have these kinds of sophisticated cognitive abilities, then those without those cognitive capabilities will be unable to entertain first-person thoughts.

Small children might satisfy this description, not yet having the capacities for conceptual spatial reasoning of the sort Evans requires for a subject to construct and maintain the cognitive map on which she locates herself. But whilst they lack the capacity for conceptual spatial reasoning, we might think that children at this stage of development have the capacity to entertain first-person thoughts, or thoughts in which a proto-self-concept is employed. It seems that Evans's account places excessive requirements on a subject's being able to think about an object.

The substance of Evans's account is that one is able to think about oneself because one can locate oneself as an object with a spatio-temporal location. Evans provides a sophisticated analysis of how this could be achieved, and the conditions one must meet for one's thought to employ the self-concept. In particular, it is necessary, according to Evans, for one to conceive of oneself as part of the objective order in order to employ the self-concept in thought. This way of thinking about oneself is enabled by the subject's engaging in certain kinds of conceptual spatial reasoning (Evans, 1982, p. 223). An alternative approach would be to accept that it is important that one's thought about oneself should be made possible by the information one receives about one's spatial location, without it being required that one possesses certain high-level capacities for spatial reasoning. We want to retain the benefits of Evans's account, but provide an account of first-person thought that allows that children can think about themselves self-consciously even whilst they lack the capacities for conceptual spatial reasoning that Evans demands. This will allow us to argue, as Evans does, that the failure to introspectively observe a self need not prevent us from offering an account of first-person thought and first-person reference. We can think of ourselves first-personally by possessing a nonconceptual point of view on our environments.

Bermúdez (1998) argues that we develop what he calls a nonconceptual point of view on the world, which marks a substantial form of self-awareness because it expresses the capacity to distinguish ourselves from our environments in spatial terms, but, at the same time, such a point of view is an antecedent to fully conceptual self-conscious thought. As such we can take Bermúdez to offer an account of the content of proto-first-person thought.

Bermúdez's argument develops from a puzzle concerning our ability to use the first-person pronoun, which has its origins in Anscombe's (1994) dissatisfaction with the self-reference rule account of first-person reference. In order to use the first-person pronoun to refer to oneself intentionally, one needs to be able to entertain first-person thoughts. Using

the first-person pronoun to refer to oneself requires that one intends to use it to refer to oneself, which means that any use of the first-person will presuppose a prior capacity for first-person thought (Bermúdez, 1998, pp. 14-21). If we expect first-person thought to be accounted for in terms of linguistic capacities, then we have a problem in explaining how it is that we come to use the first-person pronoun in the appropriate way. Bermúdez calls this the paradox of self-consciousness. In order to avoid the paradox we need to offer an account of first-person thought that does not rest on a subject's linguistic capacities: we need an alternative account of how it is that we come to entertain first-person thoughts. His suggestion is that it is by having nonconceptual first-person thoughts—thoughts that employ non-conceptual contents—that we can come to entertain thoughts that employ the first-person concept.

We can think of Bermúdez's account as providing us with a primitive self-concept. We can still think of it as one of the constituents of thought, a part of one's thought that specifies the self because it allows for the differentiation of self from environment. In a similar fashion to Evans's self-concept, though, Bermúdez's first-person proto-concept is based on one's knowledge of oneself as located in a spatial world.

Like Evans, Bermúdez requires that a subject have a 'simple theory of perception' that allows her to develop an integrated representation or cognitive map of her environment.[33] Bermúdez's account, though, is less demanding than Evans's (Bermúdez, 1998, p. 223). One becomes spatially aware through perception of oneself and one's environment, but also through grasping one's potential for acting in the world. As a minimum condition, a subject will need to map the relations between different affordances onto an integrated representation of the world.[34] When she adds to this integrated representation of the

[33] An equally important component of the non-conceptual point of view is the capacity for feature recognition allowing one, in some minimal sense, to distinguish one's experiences from the objects that they are experiences of: "The notion of a nonconceptual point of view brings together the capacity to register one's distinctness from the physical environment and various navigational capacities that manifest a degree of understanding of the spatial nature of the physical environment" (Bermúdez, 1998, p. 221). I will concentrate on the spatial aspects of Bermúdez's account here since it makes explicit the way in which he suggests that one thinks of oneself self-consciously as an element of an objective world.

[34] Bermúdez takes the notion of an affordance from J.J. Gibson's (1979) *The Ecological Approach to Visual Perception*. When I enjoy a conscious visual perception of my laptop then the information I (or rather, the visual system) receive(s) about the world in some respect concerns me. When I look at my hands typing on the keyboard I perceive myself. However even the percept of my laptop without my hands typing provides information about the self because the laptop is at a certain distance from me, in a certain direction relative to where I am, and affords certain kinds of actions on my part. I can grasp it in many ways—ways that are specific to myself, to my own abilities and strengths. Gibson writes: "If a terrestrial surface is nearly horizontal (instead of slanted), nearly flat (instead of convex or concave), and sufficiently extended (relative to the size of the animal) and if its substance is rigid

affordances within her environment the capacity to think about different routes between one place and another, and the capacity to think about places independently of the objects or features found at particular times at those places she will possess an integrated representation of the world (Bermúdez, 1998, pp. 224-7). Bermúdez does not require that a subject engage in conceptual reasoning about the locations of objects relative to each other and herself relative to those objects, but rather the calibration of affordances into a representation of the world, and the subsequent capacities which she can build up from that ability to represent her environment, can be achieved through navigation around the world as she moves from one affordance to another.

Bermúdez's non-conceptual point of view is supposed to provide subjects with the capacity for a kind of proto-thought about themselves that will be self-conscious without requiring overly sophisticated capacities for spatial reasoning. His account offers a way of retaining the idea that we think of ourselves as material objects when we entertain first-person thoughts, without being committed to the claim that subjects must engage in conceptual reasoning in order to be able to entertain those thoughts.

5.7 Thoughts about disembodiment

However, there is a further way in which we might object to an objective account—whether Evans's or Bermúdez's—on the grounds that it is too strong. We might object to the account on the grounds that it seems coherent for a subject to wonder whether or not she is embodied. An objective account of the self-concept cannot allow that such a thought is conceptually coherent. The problem is not merely that of wondering how a subject might meet the knowing-which requirement in a situation in which she has no current perceptual information about the world and her body. Rather, the question is whether we need to adapt the requirement so that a subject need not, after all, think of herself as a material body with a spatial location in order to entertain a first-person thought. The knowing-which requirement, at least as Evans conceives it to be, seems to be too strong to be a tenable condition on first-person thought. The potential coherence of the thought expressed by an utterance of the question 'Am I disembodied?' calls into question the view that our self-concept is a concept

(relative to the weight of the animal), then the surface affords support. It is a surface of support, and we call it a substratum, ground or floor. It is stand-on-able, permitting an upright position for quadrupeds and bipeds. It is therefore walk-on-able and run-over-able. It is not sink-into-able like a surface of water or a swamp, that is, not for heavy terrestrial animals. Support for water bugs is different." (J. J. Gibson, 1979, p. 127)

of a material object.

Pace Evans, it seems that we might be able to think about ourselves even if we had never had any kind of perceptual information available to us to indicate that we are material beings located in space and time. As such, we would have the capacity to entertain first-person thoughts without having any conception of ourselves as material and spatio-temporally located. The scenario might obtain if it were really possible for a brain in a vat to exist and to engage in cognitive activities without being in receipt of any (false) perceptual inputs. The brain would not be able to refer demonstratively to objects in the world, since it would have no perceptual input. It would be unable to locate itself in the world too. It would be unable even to conceive of itself as a material object, since it would fail to have any of the perceptual information necessary to acquire such a self-conception. Yet, it seems plausible that the subject would still be able to entertain some first-person thoughts (Peacocke, 2010, pp. 105-106).

Alternatively, we might think it conceivable that a subject could find herself in a situation where she no longer takes herself to be a material being. She might think she no longer has a corporeal presence, but is a purely thinking thing. Even though she remembers perceiving objects in the world and working out where she was located relative to those objects, she no longer conceives of herself as part of the objective order. If such a situation were to obtain, the subject might wonder, for example, whether she is disembodied. A subject in a particularly effective sensory deprivation tank who has been administered with local anaesthetic so that she has no proprioceptive experience of her body whatsoever might entertain this kind of thought. Evans cannot accept that such a thought is conceptually coherent: the knowing-which requirement on first-person thought means that, for him, all first-person thoughts will be thoughts in which we employ a concept of a material object. As such, it would be incoherent or nonsensical to wonder whether one is disembodied. One would be wondering whether the material being that one is, is non-material.

The coherence, then, of the thought expressed by an utterance of the sentence ‘Am I disembodied?’ is a problem for the objective account of the self-concept. We might think that because such thoughts are coherent, we must abandon the knowing-which requirement altogether, assuming in doing so that a subject can only satisfy the requirement by thinking of herself as a material object. But we might alternatively conclude that a subject who questions whether she is disembodied by entertaining a first-person thought must be meeting the knowing-which requirement in some other way. The thought is coherent because we can satisfy the knowing-which requirement not only by thinking of ourselves as bodies in space,

but also by thinking of ourselves as thinking beings. The knowing-which requirement is met not by knowledge of oneself as material object, but knowledge of oneself as a thinker.

Bill Brewer (1995) considers something similar to this in developing a Cartesian objection to Evans's account of first-person thought. Brewer suggests that the Cartesian might respond to Evans by distinguishing two different kinds of first-person thought. We might allow that some of our first-person thoughts are thoughts about a material being. The thought that my legs are crossed is a thought about a material object and it employs a concept of a material object—my body. However, the objection continues, when we entertain first-person thoughts in which we ascribe psychological properties to ourselves we do not refer to our bodies: "In self-ascribing mental properties, I refer to my essential self, the immaterial Cartesian ego" (Brewer, 1995, p. 295). Brewer characterizes the Cartesian position as follows:

"The crux of this Cartesian conception is the claim that commonsense ascriptions of both mental and physical properties to a single entity are ultimately misleading. Fundamentally, the entities figuring in mental and physical self-ascriptions are ontologically quite separate." (Brewer, 1995, p. 296)

We need not commit to such a Cartesian view of first-person thought in arguing that we can meet Evans's knowing-which requirement in different ways on different occasions. For the Cartesian, our psychological self-ascriptions are thoughts about a Cartesian ego. We need not, though, take such thoughts to refer to an immaterial self, whilst still maintaining that one can entertain first-person thoughts in which one conceives of oneself as a thinker.

The idea that we meet the knowing-which requirement in different ways—not simply by employing different kinds of perceptual information on different occasions to locate ourselves on a cognitive map of our environment, but by having knowledge of ourselves as thinkers as well as knowledge of ourselves as spatially-located material bodies—should also be distinguished from Wittgenstein's (1958) distinction between two uses of the first-person pronoun. Recall that we saw in Chapter Two that Wittgenstein takes the first-person pronoun to be used in two different ways:

"There are two different cases in the use of the word "I" (or "my") which I might call "the use as object" and "the use as subject". Examples of the first kind of use are these: "My arm is broken", "I have grown six inches", "I have a bump on my forehead", "The wind blows my hair about". Examples of the second kind are: "I see so-and-so", "I hear so-and-so", "I try to lift my arm", 'I think it will rain", "I have a toothache". One can point to the difference between these two categories by saying: The cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of an error, or as I should rather put it: The possibility of an error has been provided for." (Wittgenstein, 1958, pp. 66-7)

Wittgenstein's distinction between uses of the first-person pronoun as object and as subject seems to be a distinction between thoughts that are identification-dependent and thoughts that are identification-free. In contrast, the distinction I am advocating here is a distinction between first-person thoughts in which one conceives of oneself as a thinker and first-person thoughts in which one conceives of oneself as a bodily being.

It seems plausible that one can entertain first-person thoughts successfully without having any conception of oneself as a spatio-temporally located material object. This might be taken to signify that an objective account of the first-person concept must be rejected. Alternatively, though, we might take it to signify that a subject can meet Evans's knowing-which requirement in several ways. A subject entertaining a successful first-person thought might satisfy the knowing-which requirement either because she has the capacity to locate herself in space and time or at least think of herself as an element of the objective order, but she might also satisfy the knowing-which requirement because she has some way of conceiving of herself as a thinker.

The recognition that one should be able to conceive of oneself as a thinker rather than just as an element of the objective order is of relevance to our ability to meet the self-consciousness condition on first-person thought detailed in Section 1.2 of Chapter One. First-person thought is thought in which one knows one is thinking about oneself. In the Appendix to Chapter Seven of *The Varieties of Reference* (1982) we find the following passage:

> "Clearly it follows, from the fact that someone is thinking about himself, that the object of that thought is a thinker: specifically, the thinker of that thought. So it must follow, from the fact that someone realizes that he is thinking about himself, that he realizes that the object he is thinking about in that thought is the thinker of that thought. Hence, it is not implausible to hold that there is an essential connection between self-consciousness and the conception of oneself as the subject of certain psychological properties: specifically, thinking, and any properties that are necessarily possessed by a thinker." (Evans, 1982, p. 259)

Evans's objective account of the self-concept will have difficulties accommodating the self-consciousness of first-person thought, since it seems that an account must explain not how a subject knows which object in the world she is, but rather how a subject can know that she, a thinker, is thinking about herself. As Evans puts it, one must think of oneself as a "self-thinker" (Evans, 1982, p. 259). If we allow, though, that one conceives of oneself as a thinker, or at least that one way that one conceives of oneself is as a thinker, then we are in a better position to explain how it is that when one entertains a first-person thought one knows that one is thinking about oneself.

5.8 Conclusion

In this chapter I have examined the idea that the self-concept is a concept of a material object—an object we can know about through ordinary perceptual experience of the world and proprioceptive and kinaesthetic experience of the body from the inside. According to Evans's account of first-person thought we must think of ourselves as spatially located persons if we are to be credited with entertaining thoughts about ourselves. Our introspection-based thoughts are based on introspective awareness of mental states only, but in entertaining such thoughts we must make use of a concept of a material object—a conception of ourselves as persons with a particular spatial location. This is necessary according to Evans if we are to be able to discriminate ourselves from all other persons.

I considered several objections to this account. First I looked at the problem that the vulnerability of our perceptual apparatus to failure raises for Evans. O'Brien (2007) argues that a subject will only meet Evans's knowing-which requirement in a situation in which her perceptual apparatus and memory are seriously impaired if she has a prior, unexplained capacity for first-person thought. A second problem is that Evans's account requires too much sophistication on the part of the thinker of a first-person thought by requiring a successful thinker to be able to engage in conceptual reasoning about space. Evans requires too much from his subjects for his account to be credible. In response to this objection I presented Bermúdez's more moderate account of what I labelled a proto-first-person-concept. The most serious problem with the account, though, is that it cannot treat questions about a subject's embodiment as coherent and sensible. If we think that it is genuinely possible to question whether one is embodied, then we must either reject the objective account, or add to it the stipulation that one can also conceive of oneself as a thinker without making any commitments to the embodiment or disembodiment of that thinker.

Chapter 6: Fully self-conscious thought

6.1 Introduction

In Chapter Three I argued that, while the self-reference rule account of first-person reference seemed adequate in providing a semantic account of the first-person, we need to say more if we are to explain the self-consciousness of first-person thought. The self-reference rule determines a reference for each use of the first-person concept, but it does not seem to give us the resources to explain how it is that a subject who employs the first-person concept in thought knows that she is thinking about herself. Accounting for the self-consciousness of first-person thought was one of the conditions I placed on a successful account of this special way of thinking about ourselves in Chapter One. On this basis I looked both at a descriptive account and an objective account of first-person thought, but found both to be lacking.

In this chapter I will reconsider how we might use the self-reference rule to explain the self-consciousness of first-person thought. Without some kind of supplementation the self-reference rule account of first-person reference does not and is not intended to provide us with such an explanation. However, that doesn't mean that it might not contribute to a satisfactory account of "fully self-conscious" (Peacocke, 2010, pp. 78-79) first-person thought.

6.2 Accounting for fully self-conscious first-person thought

In Chapter One we looked at the importance of the self-consciousness of first-person thought. First-person thoughts have special cognitive significance, which is explained by these being thoughts in which one knows one is thinking about oneself. When I think about myself by entertaining a first-person thought I know that I am thinking about myself. The self-consciousness of first-person thought involves "a subject referring to herself as object, knowing that it is herself, rather than another, that she is referring to" (O'Brien, 2007, p. 64). The question, then, is this: how does one know that it is oneself that one is thinking about when one entertains a first-person thought?

We have already seen that the self-consciousness of first-person thought presents us with great difficulties in providing a satisfactory account. The perceptual information we have about ourselves does not seem to provide us with a way of thinking about ourselves that is a

genuine form of self-consciousness. Moreover, as we saw in Chapter Three, what seems like a plausible account of first-person reference—the self-reference rule account—does not seem to provide us with the resources to explain the self-consciousness of first-person thought. I will rehearse this second problem again now. We will see that whilst the self-reference rule alone does not provide us with the resources to explain how it is that one can know that one is the subject of one's first-person thought, a subject's grasp of the rule can have a role to play in the explanation of that phenomenon.

Anscombe (1994) intends for her thought experiment concerning a community that uses the term 'A' to show that the first-person cannot be a term that is used by an individual to refer to herself and herself only. She describes an imagined community who use the term 'A' to refer to themselves in the following way:

"Imagine a society in which everyone is labelled with two names. One appears on their backs and at the top of their chests, and these names, which their bearers cannot see, are various: '*B*' to '*Z*' let us say. The other, '*A*', is stamped on the inside of their wrists and is the same for everyone. In making reports on people's actions everyone uses the names on their chests or backs if he can see these names or is used to seeing them. Everyone also learns to respond to utterance of the name on his own chest and back in the sort of way and circumstances in which we tend to respond to utterance of our names. Reports on one's own actions, which one gives straight off from observation, are made using the name on the wrist. Such reports are made, not on the basis of observation alone, but also on that of inference and testimony or other information. *B*, for example, derives conclusions expressed by sentences with '*A*' as subject, from other people's statements using '*B*' as subject." (Anscombe, 1994, pp. 143-144)

Each individual within the 'A'-using community uses 'A' to speak about the individual upon whom they see the letter 'A' stamped on the inside wrist. Barring accidents then, as with the first-person pronoun, each person will use 'A' to speak about himself or herself only.[35] However, Anscombe, notes, 'A' is not the same as 'I'. 'A'-users are, by definition, not self-consciously speaking about themselves:

"In my story we have a specification of a sign as a name, the same for everyone, but used by each only to speak of himself. How does it compare with 'I'? —The first thing to note is that our description does not include self-consciousness on the part of the people who use the name 'A' as I have described it." (Anscombe, 1994, p. 144)

This is not to say that 'A'-users are not aware of themselves at all. They observe their own bodies and their own actions. They hear about themselves through the testimony of others. We might think of all these kinds of awareness as forms of self-consciousness:

[35] Anscombe allows that 'A'-use might go astray: "Of course, a man *B* may sometimes make a mistake through seeing the name '*A*' on the wrist of another, and not realizing it is the wrist of a man whose other name is after all not inaccessible to *B* in the special way in which his own name ('*B*') is." (Anscombe, 1994, p. 144)

consciousness of the individual one happens to be. But fully self-conscious thought is not merely thought in which one thinks about the individual one happens to be. Fully self-conscious thought is thought in which one knows that one is thinking about oneself. Although 'A'-users use 'A' to speak about themselves, they don't know that this is the case. Their utterances do not express self-consciousness.

We can transpose the thought experiment to the level of conceptual thought, imagining a community that uses an 'A'-concept in the same way. A subject will employ 'A' to think about the individual on whose wrist she sees the letter 'A' stamped. Again, if there are no accidents each individual will use 'A' to think about himself or herself. As with Anscombe's 'A'-using community, though, an 'A'-user in this scenario will use 'A' to think about the individual she happens to be, but in doing so she will not know that she is thinking about herself. 'A'-thoughts and first-person thoughts differ in that the latter are fully self-conscious thoughts, whilst the former are not. What, then, is the relevant difference between 'A'-thoughts and first-person thoughts that explains this difference?

Anscombe seems to take the difference to be that 'A' is a term with reflexive reference, whilst the first-person is not. Anscombe's reasoning is as follows: if 'A'-use is governed by the self-reference rule but the concept fails to be employed in fully self-conscious thought, then it cannot be the case that the first-person concept—employed in fully self-conscious first-person thought—is governed by the same rule.

Even if we accept, for the sake of argument, that the reference of 'A' is determined by the self-reference rule, the thought experiment does not show that the self-reference rule is incompatible with self-conscious first-person thought (Garrett, 1998, p. 100). If the self-reference rule account of first-person reference did not allow for first-person thought to be self-conscious thought, then we would have to abandon that approach to first-person reference. But this is not the case, and the thought experiment certainly does not show it to be the case. Neither does it show, as we will see in more detail below, that the self-reference rule cannot have a role to play in an explanation of self-consciousness.

What is more, the reference of 'A' is not determined by the self-reference rule. The reference rule for 'A' will be something like the following: a token of 'A' refers to the individual on whose wrist 'A' is inscribed (Peacocke, 2010, p. 84). This is not the same as the self-reference rule, which states that a token of the self-concept refers to the producer of the thought in which the token is employed. Although the thought experiment is such that 'A'-users will typically refer to the same individual as users of the self-concept, the 'A'-concept, unlike the self-concept, is not a device of reflexive reference (O'Brien, 1994). As

long as no one sees the 'A' stamped on the wrist of another person, then 'A', like the self-concept, will be a concept with reflexive reference. Each token of 'A' will refer to the individual who produced it. So, the 'A'-concept when used correctly and the self-concept both refer to the producer of a token of that concept. But whilst 'A'-use might be consistent with reference that is determined by the self-reference rule in this way, it is not reference that is governed by the rule. As O'Brien (1994) says of 'I'-use and 'A'-use:

"[…] a person who comprehendingly uses 'I' uses it precisely *as* a device of self-reference. The person who comprehendingly uses 'A', on the other hand, uses it as a device that refers to that thing, whatever it is, which falls under their observation in certain ways; it so happens that it is they themselves that tend to fall under their observation in those ways." (O'Brien, 1994, p. 279)

An individual who uses 'A' entertains a thought that employs a concept that happens to have reflexive reference, but the 'A' user does not know that this is the case. A concept's simply having reflexive reference will not be enough for it to be used in thoughts that are truly self-conscious. One difference, then, between first-person thoughts and 'A'-thoughts is that a subject who entertains a first-person thought has knowledge that the concept she employs in thought is a concept that has reflexive reference, whilst the thinker of the 'A'-thought does not know that this is the case. It seems that we must ascribe to the subject some kind of knowledge that the concept she employs is a concept that has reflexive reference if we want to explain the self-consciousness of first-person thought.

We cannot, however, explain the self-consciousness of first-person thought solely in terms of a subject's knowing that the concept she employs in thought has reflexive reference if this is understood simply as knowledge of the self-reference rule. A subject's grasp of the self-reference rule alone will not be enough to explain the fact that she knows that she is thinking about herself. The self-reference rule states simply that a token of the self-concept refers to the thinker of the thought in which the token is employed. But if a subject has only this third-personal understanding of reflexive-reference, this in conjunction with the awareness she has of a thought when she entertains it—say, the thought *I am tall*—will only give her the knowledge that the self-concept employed in the thought refers to the producer of that thought. She will not know that the self-concept refers to herself in the way required for the thought to count as self-conscious.

To summarize the argument so far, 'A'-thoughts seem to be different from first-person thoughts because they are not fully self-conscious: the thinker of an 'A'-thought does not know that she is the subject of her thought. 'A'-thoughts differ from first-person thoughts because users of the 'A'-concept do not know that it refers reflexively. But it cannot be mere

grasp of reflexive reference that explains the difference between self-conscious first-person thoughts and non-self-conscious 'A'-thoughts. This is because a subject's knowing that 'I' is a concept that refers to the individual who produced that thought will not allow a subject to infer that a particular token of the self-concept refers to herself. It allows her to infer only that that particular token of the self-concept refers to the person who produced it. So, if all we ascribe to the subject is a third-personal understanding of reflexive reference in the form of grasp of the self-reference rule, and the introspective awareness she has of her thought when she entertains it, then we cannot explain the self-consciousness of first-person thought.

It seems to be this line of reasoning that motivates Anscombe's suggestion that the only adequate way of understanding the self-reference rule gives us a circular account of the first-person (Anscombe, 1994, pp. 141-143). Anscombe considers whether we might merely say of the first-person that it is the word that each of us uses to speak of himself or herself. But this general reflexive rule is compatible with the subject not knowing that she is speaking about herself: it is a rule of reflexive reference, but not one that ensures that the subject will know that she is engaged in referring reflexively. To illustrate the point we might think of Oedipus as referring to himself when he utters the words 'the slayer of Laius', yet he does not know that he is engaged in referring reflexively (O'Brien, 2007, p. 56).

Anscombe considers whether we might ensure that the subject knows that she is thinking about herself by stipulating that each of us uses the word 'I' to *knowingly and intentionally* speak of himself or herself. What differentiates Oedipus's use of 'the slayer of Laius' from his use of the first-person pronoun is that while he speaks of himself when he uses 'the slayer of Laius', he does not do so knowingly and intentionally (Evans, 1982, p. 258). The problem we have though, says Anscombe, is explaining what the subject knows without re-employing the first-person. Once again transposing the discussion to the level of conceptual thought, the suggestion is that we can capture the self-consciousness of first-person thought by saying that a subject uses the self-concept to knowingly think about herself. Anscombe's response is that we must understand the 'herself' as what she calls the 'peculiar indirect reflexive': as a reflexive which can only be explained in terms of the first-person itself.

Anscombe's interests lie in an account of the reference of the first-person pronoun. She assesses the self-reference rule in the manner above because she wants an account of the meaning of 'I'. But, she also wants an account that accommodates the self-consciousness of first-person pronoun use. Since Anscombe wants both an account of first-person thought and an account of the self-consciousness of first-person thought and she expects a subject's grasp of the self-reference rule to explain both, she insists that the statement of the self-reference

rule must itself employ the first-person. We cannot give an account of the self-consciousness of first-person thought without making use of the first-person. If we have the same aspirations, then we too will run into problems of circularity, because, to account for the self-consciousness of first-person utterances solely by ascribing to the subject grasp of the rule that determines the reference of 'I', we will be forced to specify that rule in first-person terms. This, of course, gives a circular statement of the reference rule, and we are left without an account of first-person thought. Hence Anscombe says:

"If that is right, the explanation of the word 'I' as "the word which each of us uses to speak of himself" is hardly an explanation!—At least, it is no explanation it that reflexive has in turn to be explained in terms of 'I';" (Anscombe, 1994, p. 142)

As we saw in Chapter Three, as a semantic account, the self-reference rule account is perfectly adequate. As long as we avoid the first-personal version of the rule—'I' refers to me, perhaps—then we have a satisfactory account of the reference of the first-person pronoun, and I suggest first-person thought also. But can we give an account of the self-consciousness of first-person thought without ascribing to the subject a grasp of the first-person version of the rule?

What Anscombe's discussion illustrates is that we should not expect an account of first-person reference to explain, by itself, the self-consciousness of first-person thought. Our account of first-person thought cannot be the only place to look for an account of the self-consciousness of those thoughts. This does not, of course, mean that our account of the reference of the first-person cannot have a role to play in an explanation of the self-consciousness of first-person thought. Anscombe's troubles arise because she seems to assume that we must provide an account of first-person reference that also explains the self-consciousness of first-person thought. In particular she seems to think that we must explain both phenomena by ascribing to the subject grasp of the self-reference rule only.

6.3 Peacocke's account of fully self-conscious thought

If we think that our explanation of the self-consciousness of first-person thought must draw only on the subject's awareness of her first-person thought as she entertains it, and her grasp of the self-reference rule, then we seem to be obliged to ascribe to the subject grasp of something more than a third-person reflexive rule of reference. Rather than adapting the rule that a subject takes to determine the reference of the self-concept, though, we might add something else to our explanation.

According to O'Brien (1994), Anscombe's thought experiment concerning the 'A'-using community indicates precisely the need for us to add something to our account if we are to explain the self-consciousness of first-person thought. What we need, as well as the subject's awareness of her first-person thought and her knowledge that a token of the self-concept refers to the producer of the thought in which the token is employed, is for the subject to know that she is engaged in producing a token of the self-concept (O'Brien, 1994, p. 280).[36]

Christopher Peacocke (2010) also suggests that we can explain the self-consciousness of first-person thought by supplementing the ascriptions we make to the subject. In addition to the awareness she has of her first-person thought when she entertains it and her knowledge of the self-reference rule, we need to ascribe to the subject the knowledge that she is producing that first-person thought. If we allow that a subject knows that she is producing a token of the self-concept, then we can explain how she can know that the self-concept in her thought refers to herself by ascribing to her nothing more than a grasp of the self-reference rule.

Peacocke gives the following explanation of how a subject can come to know that she is the subject of her first-person thought:

"Judging something is a mental action, of which a thinker has distinctive action-awareness. So we start from the point that our normal thinker has an awareness of his mental action, an action-awareness with the content:

(1) I am judging that I am F.

The circumstances of our case are ordinary, and the thinker takes the content of this awareness at face value. That is, he not merely has an awareness as of (1)'s being the case, he endorses the content of the awareness, and judges (1). Our thinker also knows the fundamental reference rule for the first-person concept. That is, he knows

(2) Any use of *I* in a thinking refers to the thinker of that thinking.

From this knowledge of (1) and (2), and a presumed background knowledge that judging a content is a form of thinking, our thinker knows

(3) *I* in this judging *I am F* refers to me.

That is, since "me" is just the accusative form of the first-person, our thinker knows:

(4) I have the property λx[*x*'s use of *I* in this judging refers to *x*]." (Peacocke, 2010, p. 81)

Let's consider a situation in which the thought entertained by the subject has the content *I am tall*.[37] The thought entertained must employ the self-concept, otherwise it would not be

[36] I will look briefly at O'Brien's (2007) view of what it is that we must add to a subject's grasp of the self-reference rule in order to explain the self-consciousness of first-person thought in Section 6.4.

[37] Peacocke's account is based around judgement: it is important that the subject is engaged in active judgement rather than merely having a passive thought occur to her, otherwise we could not ascribe to the subject action-awareness which has an essential role to play in the account under consideration. Peacocke does allow that we can ordinarily know that we are thinking about ourselves when we entertain passive first-person thoughts, first-person thoughts which simply occur to us. We are entitled

a self-conscious first-person thought. In the scenario being considered the subject self-ascribes the property of being tall. If the first-person thought is to be a self-conscious thought, then the subject must know that the self-concept employed in her thought *I am tall* refers to herself: *I* in *this* thinking *I am tall* refers to me.[38]

How can we explain how the subject knows this? If we ascribe to the subject knowledge of the self-reference rule—that a token of the self-concept refers to the producer of the thought in which the token is employed—then what else is needed to explain her knowledge that it is to herself that the self-concept refers? We saw earlier that if we assume that the subject has nothing but an awareness of the thought that she entertains, then grasp of the self-reference rule will only be sufficient to give her knowledge that the self-concept refers to the thinker of the thought entertained. We must, then, ascribe to the subject knowledge that the thought is one of her own: that it is she that is the producer of the thought. Peacocke explains this knowledge in terms of a subject's having action-awareness with the following content: *I am judging I am tall*. He suggests that the subject can take the content of this action-awareness that she has of what it is that she is doing 'at face value', and endorse the content by judging *I am judging I am tall*.

According to Peacocke we can explain the self-consciousness of first-person thought with reference only to a subject's grasp of the self-reference rule and "the presence of other facts common to any ordinary case of thinking" (Peacocke, 2010, p. 81). What Peacocke is proposing is that subjects do not just have introspective awareness of their mental actions when they perform them: one stands to one's agentive thoughts as agent and one is aware of oneself as producing those thoughts. We can, then, use a subject's awareness of her first-person agentive relation to her thoughts to explain how she can come to know that she is thinking about herself.

---

to this knowledge, he says, because we can rely on it being the case that we ordinarily produce these passive thoughts, even though we have no awareness of producing them (Peacocke, 2010, pp. 90-91).

[38] Peacocke follows Evans (1982, p. 258) in thinking that to explain how a subject knows she is thinking about herself what we must show is that the subject knows she has the property of being a self-referrer. Evans makes this claim in response to Anscombe. It is not clear to me that Anscombe is guilty of the mistake that Evans and Peacocke attribute to her, the mistake of claiming that to explain the self-consciousness of first-person thought we must ascribe to the subject a *self-conscious* first-person intention. Rather, I think Anscombe's dissatisfaction with the self-reference rule account of first-person reference is as I have described above: that we cannot give a satisfactory account of first-person reference which also explains the self-consciousness of first-person thought since we must ascribe to the subject a first-person intention. I find nothing in Anscombe to support the idea that she thinks that the indirect reflexive we must use in the self-reference rule is a *self-conscious* 'herself'. Nevertheless, I think Evans is correct in his assessment of what it is that we need to offer in an explanation of the self-consciousness of first-person thought: an explanation of how it is that a subject can know that she has the property of being a self-referrer.

If we ascribe to the subject the thought *I am judging I am tall*, along with a grasp of the self-reference rule, then the subject will be in a position to know that the token of the self-concept employed in her thought refers to herself, which is what is required if the subject is to know that she is the subject of her first-person thought. Peacocke explains the subject's being able to entertain the thought *I am judging I am tall* in terms of her having action-awareness with that content, and taking that content at face value. What he does is ascribe to the subject more than an awareness of the content of the thought that she entertains. He also ascribes to the subject the awareness that she is engaged in producing that thought. This is action-awareness, the content of which he characterizes simply as *I am judging I am tall*. By taking the content of that awareness at face value, Peacocke says, the subject can judge *I am judging I am tall*. The judgement in combination with her knowledge of the self-reference rule enables the subject to infer that the self-concept in her thought *I am tall* refers to herself. In knowing this she knows of her first-person thought *I am tall*, that she is thinking about herself, and so we have an account of fully self-conscious first-person thought.

We can, then, understand why thinkers of 'A'-thoughts lack the knowledge that they are referring to themselves in terms of the disparity between their action-awareness, if we allow that they have any action-awareness at all, and the reference rule that governs use of the 'A'-concept. If we allow for the sake of argument that an 'A'-user will have action-awareness, that action-awareness will have the first-person content *I am judging A is F*. Together with her knowledge of the fundamental reference rule for A, which states that any use of A refers to the person on whose inside wrist 'A' is inscribed, this only allows the subject to infer that *A in this judging* A is F *refers to the person on whose wrist that 'A' is inscribed* (Peacocke, 2010, pp. 83-84). This, of course, falls short of knowledge that one is referring to oneself. The matter would be different if we could ascribe to the subject action-awareness with a different kind of content:

> "If the content of the action-awareness involved in conscious judgement were not merely of the first-person form I am judging so-and-so, but were rather I, the person on whose wrist 'A' is inscribed, am judging so-and-so, it would be possible to close the gap in the argument. But it is not. The fact that action-awareness has a first-person content is essential to the argument" (Peacocke, 2010, pp. 84-85)

All the 'A'-user can know is that she is an 'A'-referrer. To be in a position to conclude that she is a self-referrer she would also need to know that she is the person on whose wrist 'A' is inscribed. The difference between thinkers of 'A'-thoughts and thinkers of first-person thoughts lies in the fact that 'A' is not governed by the self-reference rule. For the sake of argument we can allow that thinkers of 'A'-thoughts even have the same kind of action-

awareness as ordinary thinkers. Because action-awareness has first-person content it will not entitle a thinker of an 'A'-thought to conclude that she is thinking about herself. The thinker of the 'A'-thought will know that her thought *A is tall* is one of her own active thoughts, a thought that she has produced, but because the reference rule governing 'A'-use does not tell us anything about the relation between the producer of the thought and the reference of the concept, then this knowledge will not help her to reach the conclusion that she is thinking about herself when she employs the 'A' concept.

Peacocke's account might seem to be open to the objection that it engages us in a regress. If we explain the self-consciousness of first-person thought by ascribing to the subject a second first-person thought, then we seem to face a regress. We ascribe to the subject a second first-person thought in order to explain her knowledge that she is referring to herself when she entertains her first first-person thought. But then, it might be objected, we must explain the self-consciousness of the second first-person thought that is used to explain the subject's knowledge that she is a self-referrer.

First-person thoughts are self-conscious thoughts. If we explain the self-consciousness of one particular first-person thought *I am tall* in terms of a subject's entertaining another first-person thought, then we also have to offer an explanation of the self-consciousness of this second first-person thought. If the self-consciousness of one particular first-person thought is explained by ascribing to the subject a second first-person thought, then the self-consciousness of this second first-person thought must itself be explained. We must ascribe to the subject a third first-person thought to explain the self-consciousness of the second first-person thought, but once again we have a first-person thought the self-consciousness of which must be explained, and so on.

I think it is right to take Peacocke's account of a subject's knowledge that she is referring to herself when she entertains a first-person thought to engage us in something of a regress. However, it would be a mistake to think that this is a problematic regress. It is certainly a consequence of Peacocke's account that we ascribe to the subject first-person thoughts, the self-consciousness of which is not explained by the resources immediately available. To explain the self-consciousness of these first-person thoughts we must ascribe to the subject further states of action-awareness and corresponding first-person judgements that are made simply by endorsing the content of those experiences. However, I think there is nothing problematic in this regress of explanations.

6.4 Two objections of circularity

If we allow Peacocke to ascribe to the subject the judgement *I am judging I am tall* together with a grasp of the self-reference rule, then it is clear that the subject will be able to infer *the self-concept in this thought* [I am tall] *refers to me*, and thus can be said to know that she is thinking about herself when she entertains the first-person thought *I am tall*. But can we really ascribe to the subject the judgement *I am judging I am tall*? There are several objections that might be made against this claim.

O'Brien (2007) objects to the first-person characterization of the knowledge that one is the producer of a token of the self-concept. According to O'Brien, if we characterize a subject's knowledge that she is referring to herself in first-person terms—for example, by ascribing to her the thought *I am judging I am tall*—then we face circularity or regress. We must instead ascribe to the subject "a form of self-consciousness that is prior to, and independent of, our capacity for first-person reference" (O'Brien, 2007, p. 73).

According to what she calls the two-tier strategy we can account for the self-consciousness of first-person reference simply by ascribing to the subject a first-person thought to the effect that she is producing a token of the first-person (2007, pp. 65-72). When we are interested in the reference of the first-person *pronoun*, we will ascribe to the subject a first-person thought to the effect that she is producing an utterance of the first-person pronoun. If we are interested in the reference of the *self-concept*, we will ascribe to the subject an additional first-person thought to the effect that she is producing a token of the first-person concept.[39] In this respect, O'Brien's two-tier strategy is similar to Peacocke's approach (2010): if we ascribe to the subject both a grasp of the self-reference rule and the knowledge that she is producing a token of the self-concept, then we will have the resources to explain the subject's knowledge that she is the subject of her first-person thought.

What, then, leads O'Brien to the conclusion that the account will be circular unless we specify that the subject's knowledge that she is producing a token of the self-concept be both non-conceptual and non-first-personal? O'Brien's conclusion depends on her taking the two-tier strategy to be an account of first-person reference, understood to be self-conscious self-reference, rather than merely an account of a subject's knowledge that she is referring to herself. O'Brien's reasoning, then, is similar to Anscombe's. According to Anscombe, to explain how a subject knows that she is referring to herself when she uses 'I' we must

[39] O'Brien also considers the possibility that we explain the self-consciousness of first-person thought by ascribing to the subject a mental state with non-conceptual first-person content (2007, p. 71). She argues that this alternative approach to the two-tier strategy will face the same problems as the conceptual version.

ascribe to the subject a grasp of a first-personal self-reference rule. As an account of the determination of the reference of the first-person this would be an objectionably circular statement of the reference rule for 'I'. So too, it is as an account of first-person reference that the two-tier strategy faces circularity because the approach requires that we ascribe to the subject the very thing that is supposed to be being explained: a first-person thought. The reason that there is a problem of circularity, then, is that O'Brien thinks that an account of first-person reference must itself explain the self-consciousness of first-person thought.

Why does O'Brien insist that an adequate account of first-person thought must itself explain the self-consciousness of first-person thought? A thought will be of the first-person type, she says, only if the subject knows that she is thinking about herself:

> "In essence, the problem is that any attempt to explain first-person reference as 'reflexive' reference runs into trouble, because reflexive reference can only be first-person reference if one knows that one is referring to oneself. However, that knowledge then also needs explication. It can seem obvious however that knowing that one is referring to oneself involves referring to oneself first-personally. But if that is so it seems one cannot give a non-circular account of first-person reference." (O'Brien, 2007, p. 8)

The problem is supposed to be that there may be cases of self-reference where the concept or term used is governed by the self-reference rule, but which are not cases of first-person reference. What is missing, says O'Brien, is the subject's knowledge that she is referring to herself, and therefore we must build into our account of first-person thought an explanation of how a subject knows that she is referring to herself. Yet, as Johannes Roessler (2009, p. 230) points out, it is in fact difficult to find instances of thought where the reference of the concept employed in the thought is determined by the self-reference rule, but about which the subject does not know that she is referring to herself. The cases that O'Brien takes to show the need for us to ascribe to the subject knowledge that she refers to herself in order to individuate genuine first-person thought are hard to come by. It is certainly possible to think of cases in which a subject refers to herself accidentally, for example when Oedipus uses the description 'the slayer of Laius' to refer to the person he happens to be. But these forms of accidental self-reference are not cases in which reference is determined by the self-reference rule.

Roessler distinguishes O'Brien's discussion of David Lewis's (1983a) example of two gods as providing the strongest argument for her distinction between reference governed by the self-reference rule and genuine first-person reference (Roessler, 2009, p 231). I described the example in Chapter Three, in which I accepted, temporarily, O'Brien's analysis of the two-tier strategy. In Lewis's example there are two gods, one of whom lives on the tallest mountain and throws down manna, whilst the other lives on the coldest mountain and throws

down thunderbolts (O'Brien, 2007, p. 58). We ascribe to these gods knowledge of the self-reference rule and the capacity to produce first-person thoughts. Yet, says O'Brien, the two gods are not self-conscious: neither knows which god he is. Even though they are omniscient with respect to propositional knowledge, they have no indexical knowledge. O'Brien takes the example to illustrate that an account of first-person thought must itself explain the self-consciousness of first-person thought. It is not enough merely to say that the reference of the self-concept is determined by the self-reference rule. We need to militate against the possibility of an individual's being in the position of one of Lewis's gods by insisting that an account of first-person thought explains the self-consciousness of first-person thought. When we apply Peacocke's account of how a subject knows she is referring to herself—the two-tier strategy—we run into the problem of circularity since we will be ascribing to the subject a first-person thought as part of our explanation of her capacity for first-person thought.

What O'Brien must do, in order to show that an account of first-person thought must itself also explain a subject's knowledge that she is referring to herself when she entertains a first-person thought, is show that it is possible for a subject to be in a position to produce first-person thoughts yet, as with Lewis's gods, not know that she is referring to herself. It would have to be the case that a subject could entertain the thought *I am tall*, say, and yet not be in a position to know that she has the property of being a self-referrer. According to Peacocke's account, one knows that one has the property of being a self-referrer when one can judge *I in this thinking* [I am tall] *refers to me*. A subject is able to make this judgement when she both knows the self-reference rule and judges *I am judging I am tall*. For O'Brien to distinguish reflexive-reference from self-conscious reflexive-reference in a way that forces us to build into our account of first-person reference an explanation of the self-consciousness of first-person thought, there must be a case that undermines Peacocke's account.

The type of case that would undermine Peacocke's account would involve a subject who is not able to judge *I am judging I am tall* when she judges *I am tall*. Perhaps she would have an awareness of her judging *I am tall* that would allow her to judge *there is judging I am tall*. This would not be sufficient for that subject to know that she is a self-referrer because all that a subject who judges *there is judging I am tall* and who knows the self-reference rule can infer is that *I in this judging* [I am tall] *refers to the producer of that judgement*. This is the position that Lewis's gods seems to be in: estranged from their thoughts such that they are only able to judge *there is judging I am tall*.

O'Brien's objection then, like Anscombe's worry about the adequacy of a third-person statement of the self-reference rule, is based on the claim that a subject might entertain a first-person thought yet be unable to self-ascribe the thinking of that first-person thought and thus also fail to know that the self-concept employed in her initial first-person thought refers to herself. O'Brien's sceptical challenge holds that it will be possible for a subject who judges *I am tall* to be entitled to judge, at most, something like *there is judging I am tall*.

However, Peacocke's view of the nature of the mind and the possession conditions for mental action concepts renders this outcome impossible. In particular, a subject who has action-awareness when she judges *I am tall* will not have the content *there is judging I am tall* available to her unless she also already has reason to judge *I am judging I am tall*. Action-awareness together with grasp of the concept of judgement make rational the move from judging *I am tall* to the judgement *I am judging I am tall*.

There are, I think, two ways of interpreting the sceptical challenge raised by O'Brien's worry. On the one hand, and this seems to be the challenge that Lewis's gods are supposed to raise, the suggestion is that a subject who judges that *P* will fail to have any awareness of judging that *P*, so that she cannot make the judgement *I am judging that P*. On the other, it might be suggested that, even though the subject is aware of what she is doing, she will not be able to judge *I am judging that P*.

If the sceptical challenge is merely the suggestion that a subject cannot be aware of what she is doing, then O'Brien's own commentary on the plausibility of there being an individual who could entertain thoughts yet who would have no knowledge of doing so provides us with the response we need to block the challenge:

"[…] notice that taking the case this way hinges on taking the gods to be in some peculiar way estranged from their own utterances—they do not know, as we do, that they, themselves are uttering when they are. In fact this estrangement seems to me to rule out the possibility that the gods are genuine speakers or thinkers." (O'Brien, 2007, p. 58)

The two gods appear to be estranged from their own thoughts in a way that renders them not to be genuine thinkers at all. By contrast, the way that ordinary thinkers come to know about their mental actions is, says Peacocke, through action-awareness. Certainly, the two gods, who have no first-person awareness of their mental actions, cannot come to know that they are referring to themselves when they entertain first-person thoughts. The fact that they do not means that we do not regard them as genuine thinkers at all.

However, there is a further sceptical worry that we can take O'Brien to raise. Even if a subject has action-awareness, might it be possible for a subject who judges that *P* and has action-awareness of judging that *P* to have only the content *there is judging that P* available

to her? Again, this revision of O'Brien's sceptical challenge can be blocked, this time by Peacocke's analysis of the possession conditions for the concept of judgement. Here is why. According to Peacocke, it is essential to grasping the concept of judgement that one appreciates that action-awareness of judging that *P* gives one good reason to accept *I am judging that P* (Peacocke, 2010, pp. 262-263). In other words, there is a clause in the possession conditions of the concept of judgement dealing with first-person ascriptions:

"The possession condition for concepts of mental actions contains a clause about first-person present-tense ascription that says that the thinker has reason for making such ascriptions in the presence of suitable apparent action-awareness." (Peacocke, 2010, p. 266)

A subject will not have the concept of judgement unless she is able to move rationally from action-awareness that she is judging that *P* to the self-ascription of a judgement that *P*. What this means is that a subject who judges that *P* will not be able to judge *there is judging that P*, but be unable also to self-ascribe the judgement that *P*. If she cannot do the latter, then she will not have the concept of judging, and the content *there is judging that P* will not be available to her. O'Brien's sceptical challenge, then, does not represent a genuine possibility. Either the subject has a grasp of the concept of judgement, with the consequence that when she has action-awareness of judging that *P*, then she will appreciate that she has reason to self-ascribe the judgement that *P*, or else she will not have a grasp of the concept of judgement, in which case she will not be able to judge *I am judging that P*, but neither will she be able to judge *there is judging that P* (Campbell, Forthcoming, p. 4).

If Peacocke is right, then, O'Brien's concerns over the adequacy of an account of first-person thought that does not itself explain the self-consciousness of first-person thought cannot get under way. For any ordinary judgement *I am tall*, the thinker of that thought will have an action-awareness with the content *I am judging I am tall*: she will know that the first-person thought is one of her own thoughts, that she is the *producer* of that thought. Moreover, her grasp of the concept of judgement will mean that she knows that this action-awareness gives her reason to *self-ascribe* the judgement *I am tall*. She will judge *I am judging I am tall*. Along with her knowledge of the self-reference rule, this is all that is needed for the subject to know that she is referring to herself when she entertains a first-person thought.[40]

[40] We do not need to show that a subject has actually engaged in the inference to the conclusion that she has the property of being a self-referrer. Presumably it is not the case that for every first-person thought one entertains one actually makes the inference to the conclusion that one has the property of being a self-referrer. However, an ordinary subject who has a grasp of the self-reference rule will always be in a position to know that she has the property of being a self-referrer because of the relation she stands in to her own thoughts.

In summary, for Peacocke's account to be susceptible to O'Brien's objection of circularity it must be the case that an account of first-person thought must itself explain the self-consciousness of first-person thought. O'Brien's grounds for thinking that this is required are based on the idea that a certain kind of sceptical case is possible: that a subject might be able to entertain a first-person thought *I am tall*, yet be unaware that in doing so she is referring to herself because all she is able to judge is the third-personal *there is judging I am tall*. Peacocke's account of action-awareness and the possession conditions he gives for the concept of judgement block this from being a possibility. We need not, therefore, worry about providing an account of first-person thought that explains how a subject knows she is referring to herself when she entertains a first-person thought. As such, we face no threat of circularity in explaining a subject's knowledge that she is referring to herself by ascribing to that subject an additional first-person thought.

There is, however, a further objection that Peacocke's account faces. We can question whether we can really ascribe to the subject the judgement *I am judging I am tall* without presupposing the very thing that we are trying to explain, that the subject knows that she is a self-referrer. This is different from O'Brien's concern about circularity: O'Brien's worry about circularity is the worry that we can only explain a subject's having the capacity for first-person thought by ascribing to that subject a prior capacity for first-person thought. This second worry over circularity is the worry that Peacocke's account of the subject's knowledge that she is the subject of her first-person thought presupposes that the subject knows that she is referring to herself when she entertains a first-person thought.

Bermúdez (2011) questions whether Peacocke is guilty of providing a circular explanation thus:

> "If a thinker judges *I am judging that I am F* then she must understand that she, as judger, is the author of a judgement to the effect that she herself is making a judgement about herself. What more could be required for knowledge that one is a self-referrer?" (Bermúdez, 2011, p. 1279)

What Bermúdez suggests is that just entertaining the thought *I am judging I am tall* will require that a subject knows that she is a self-referrer. That is, the thought *I am judging I am tall* cannot play a role in an explanation of the knowledge that one is the object of the thought *I am tall* since it already expresses one's knowledge that that is the case. We can see Bermúdez's point by contrasting the thought *I am judging I am tall* with the thought *Alisa is judging I am tall*. If I entertain the latter we can allow for the possibility that I do not know that it is me that is thinking *I am tall*, say, if I have forgotten my name. What seems to set the former thought apart from the latter is that I entertain the former thought—the thought *I am*

judging I am tall—specifically when I know that I am a self-referrer. The repeated use of the self-concept means that my thinking the thought *I am judging I am tall* expresses my knowledge that when entertaining the thought *I am tall* I am referring to myself. What Bermúdez queries is whether the judgement *I am thinking I am tall* can really be a premise in the subject's inference to the conclusion that she is referring to herself, since he thinks that the judgement *I am judging I am tall* already expresses this knowledge. We have not explained the self-consciousness of first-person thought if our explanation presupposes just what it is that we are trying to explain. We have not explained how a subject knows that she is referring to herself if we simply ascribe to her the knowledge that she is referring to herself.

Bermúdez's objection of circularity depends on the idea that the judgement *I am judging I am tall* is not a judgement that a subject can make without recognising already that she has the property of being a self-referrer. However, according to Peacocke a subject who judges *I am tall* will not only have action-awareness with the content *I am judging I am tall*, she will be entitled to take that content at face value and judge *I am judging I am tall*. It is when a subject makes the latter judgement simply by taking the content of her action-awareness at face value that she does not express knowledge that she is a self-referrer.

As we saw above in the response to O'Brien's objection of circularity, Peacocke's account of the possession conditions for the concept of judgement includes a clause dealing with first-person ascriptions. According to Peacocke, a subject who has a grasp of the concept of judgement will have at least tacit knowledge that her action-awareness of judging *I am tall* gives her reason to make a self-ascription of judging *I am tall*. What she appreciates is her entitlement to take the content of her action-awareness at face value.

Peacocke's account of representationally-dependent uses of the first-person concept in *Being Known* (1999) gives more substance to the idea of a subject taking the content of action-awareness at face value:

"We can say that a use of the first person, in a particular belief with the content 'I am *F*', is *representationally dependent* if

(i) 'I am *F*' is the content of one of the thinker's current mental states, a state which represents that content as correct; and

(ii) the thinker forms the belief 'I am *F*' by taking the mental state mentioned in (i) at face value, in respect of this content." (1999, p. 265)

The example Peacocke gives of this type of use of the self-concept is the following. One has a visual experience of a door being ahead of one, and for that reason forms the belief *I am in front of a door*. According to Peacocke having the experience provides the subject with a reason to entertain a thought with the same content: "The way the visual experience

represents the world as being is one which justifies his acceptance of the first-person content endorsed in his belief 'I am in front of a door'" (Peacocke, 1999, p. 264). And, even though for Peacocke the content of experience will be non-conceptual, this does not get in the way of the subject's having a reason to entertain a first-person thought in which the first-person *concept* is employed on the basis of having an experience with *non-conceptual* first-person content. If this is the case, says Peacocke, the first-person non-conceptual content of the experience stands in an implicational relation to the content of the first-person thought (Peacocke, 1999, p. 264).

Peacocke points out that representational dependence is not limited to first-person thought: the phenomenon is more general. In any situation in which one can endorse the content of experience by entertaining a thought with the same content, then one's thought will be representationally dependent on the content of that experience. This will be the case for many of our thoughts about the world based on perceptual experience. Representationally dependent uses of the self-concept based on action-awareness are just a special case of representation-dependent uses of the self-concept, and representationally dependent uses of the self-concept are just a special case of representationally dependent thoughts in general. We should note that action-awareness is not supposed to be a form of perceptual experience. Nevertheless it is "a real state of consciousness, available for rationalizing certain judgements" (Peacocke, 2010, p. 259).

Take the judgement *it is raining*. When a subject makes this judgement she will have action-awareness with content to the effect that she is judging *it is raining*. Taking the content of action-awareness at face value, she judges *I am judging it is raining*. The same holds for first-person judgements. When a subject makes the judgement *I am tall*, she will have action-awareness with content to the effect that she is judging *I am tall*. Taking the content of her action-awareness at face value she judges *I am judging I am tall*. The content of any self-ascriptive judgement based on action-awareness of making a first-person judgement will employ the self-concept twice. If the subject were not employing two representationally dependent uses of the self-concept when she entertains the thought *I am judging I am tall*, then in entertaining that thought she would, as Bermúdez suggests, express her knowledge that she is a self-referrer. However, when that judgement is made on the basis of taking the content of action-awareness at face value, this double employment of the self-concept in the judgement does not express the understanding by the subject that she herself is making a judgement about herself. Rather, it merely expresses the subject's knowledge that the judgement *I am tall* is one of her judgements (Bermúdez, 2011, p. 1278). As

Bermúdez suggests later, Peacocke's view of the content of action-awareness gives "one access to a self-referring content without requiring one to know, or even to be able to formulate in thought, the fact that one is a self-referrer" (Bermúdez, 2011, p. 1279).

We can see, then, that Bermúdez's objection of circularity depends on claiming either that we do not have action-awareness with first-person content such that there is no first-person content to take at face value and endorse in a judgement, or that we are not able to take the content of our action-awareness at face value and self-ascribe the action we are aware of performing. Either way, Bermúdez's claim is that a subject will not be in a position to judge *I am judging I am tall* when she judges *I am tall* without recognizing already that she is a self-referrer.

As we have seen, though, Peacocke's account of how a subject knows about her mental actions and his account of the possession conditions for the concept of judgement are such that a subject will both have the content *I am judging I am tall* available to her simply through action-awareness, and she will appreciate that she is entitled to endorse this content in judgement. It is in the nature of mental action that one is aware of one's being the producer of a particular mental action. We do not stand in the kind of third-person relation to the thoughts we produce that would result in awareness of thinking those thoughts with third-person content. We are always, ordinarily, aware of our actions as our own actions. It just could not be the case that one would have an awareness of making a judgement with anything other than first-person content. Moreover, a subject who has a grasp of the concept of judgement understands that she is entitled to take the content of her action-awareness of judging at face value. Thus the ascription of the thought *I am judging I am tall* to the subject, when she entertains that thought simply by taking the content of action-awareness at face value, can play a role in our explanation of how a subject comes to know that she is referring to herself when she thinks *I am tall*, rather than presupposing that knowledge.

Peacocke's account, then, gives the following three conditions on a subject's knowing that she is the subject of her first-person thought. First, she must have action-awareness with first-person content. Second, she must be able to make a first-person judgement about what she is doing simply by taking the content of that action-awareness at face value. Finally, she must have a grasp of the self-reference rule that a token of the self-concept refers to the producer of the thought in which the token is employed. If we allow that a subject can meet these three conditions, then it seems that Peacocke has provided an account that can explain the self-consciousness of first-person thought.

6.5 Peacocke's account of non-conceptual first-person content

Peacocke manages to avoid the problems of circularity described above by claiming that we have action-awareness when we entertain active thoughts such as judgements and that this action-awareness has first-person content. Crucially, he also claims that a thinker is able to take the content of that action-awareness at face value in order to make a judgement with the same content. We have already seen that a subject's grasp of the concept of judgement will have the consequence that she appreciates that her action-awareness of judging *I am tall* gives her reason to make a self-ascriptive judgement that she is judging *I am tall*. We have also seen that the idea of representation-dependent uses of the self-concept gives some substance to the idea of taking the content of an experience or state at face value. In this section I will look more closely at the claim that action-awareness has first-person content.

Peacocke does not say much about the first-person content of action-awareness in *Truly Understood* (2010). He characterizes the content of action-awareness that is essential to his account merely as *I am judging that I am F* (2010, p. 81). He adds to this the comment that, "The fact that action-awareness has a first-person content is essential to the argument […], and is not required to have more than a first-person content, in respect of whom it represents as being the agent, for the reasoning […] to go through" (Peacocke, 2010, pp. 84-85).

But we get a fuller account of this form of self-representation, as well as a better understanding of how it can play the role required in Peacocke's explanation of our knowledge that we are the subjects of our first-person thoughts, in 'Subjects and Consciousness' (2012b). Peacocke gives a constitutive account of self-representation: a subject will have a nonconceptual self-representation provided she can have mental states with the property of subject-reflexivity, and the predicative content of those subject-reflexive states is integrated in a subject's object-file on itself (Peacocke, 2012b, p. 85). Non-conceptual self-representations are linked to subject-reflexive mental states and a subject's file on itself, in such a way that having subject-reflexive mental states and a subject-file in which relevant predicative contents are integrated are necessary conditions on a subject's having non-conceptual self-representations. Both these conditions need explaining.

First, though, Peacocke individuates non-conceptual contents as of the first-person type in the following way. A non-conceptual content will be of the first-person type if it refers *de jure* to the subject of the mental state or event of which it is the content. It is the case both that "The way in which the subject is given in the awareness entirely settles that it is the subject of the awareness that the content concerns" (Peacocke, 2012b, p. 75) and that "the

determination of the subject as the reference does not proceed via satisfaction of some other condition that is involved in the nonconceptual content" (Peacocke, 2012b, p. 76).

The fact that it refers *de jure* to the subject of a mental state is to be explained by what Peacocke calls the subject-constitutive hypothesis:

"What makes a component of nonconceptual content something of the first-person type is that the fundamental condition for an instance of the type to refer to something, when it is in the content of a mental state or event M, is simply that it refers to the subject of M (M's owner)." (Peacocke, 2012b, p. 77)

We should not, therefore, characterize the first-person non-conceptual content as the structured descriptive-demonstrative content 'the subject of this experience'. The non-conceptual first-person content is unstructured. It is for this reason that it refers *de jure* rather than *de facto* to the subject of the state of which it is the content. It would be correct to say that, like the first-person content, the structured content 'the subject of this experience' refers to the subject of the state of which it is the content. Yet it is not the case, says Peacocke, that the determination of the subject as the reference of the first-person content is mediated by the determination of the reference of a descriptive component such as 'the subject of'. It is for this reason that he says it refers to the subject of the state in question *de jure* (Peacocke, 2012b, pp. 75-77).

Peacocke adds that it is a consequence of this account that when one entertains a state with non-conceptual first-person contents, then one will have the property of being a self-representer. It will be the case that one satisfies the property λy [y represents something about y] (Peacocke, 2012b, p. 77). This, of course, is to be distinguished from a subject knowing that she has the property of being a self-representer. When one has an experience with first-person content one has the property of being a self-representer, but one does not need to entertain a thought to this effect. So too, at the conceptual level, one can entertain a thought which employs the self-concept without already knowing that one is a self-representer. However, as we have seen, when one entertains a first-person thought one will always be in a position to know that one has the property of being a self-representer since one's awareness of one's thought is awareness of it as one's own.

Peacocke then spells out what is required for a subject to meet the two conditions on having non-conceptual self-representation in the following way.

The first condition is that of the state with non-conceptual first-person content having the property of subject-reflexivity. According to Peacocke a state will have the property of subject-reflexivity if it stands in the relation of subject-reflexivity to the subject of the state, which means that subject and state stand in the following relation to one another:

"λx λy [x is the subject of y & the content of y refers to x]" (Peacocke, 2012b, p. 78)

If the subject and state stand in this relation, then it will be the case that that state has content that is 'intrinsically subject-referring' or first-personal. It is a condition, then, on a state's having first-person content that the content of that state refers to the subject of the state.

The second condition is that the predicative component of the content of subject-reflexive states must be integrated into a subject's file on itself. We need to posit the integration of the information that a subject has about herself if we are to explain the synchronic and diachronic unity of self-awareness at the non-conceptual level. As Peacocke points out, the first-person concept has a role to play in the integration of conceptual information about oneself. If I entertain two first-person thoughts about myself, I am able to infer that both predicates hold of myself: I am both F and G. But we need some explanation of how I am aware of a variety of different facts about myself in a similar way at the non-conceptual level. We explain this by means of the integration of the representations of the properties that we experience ourselves instantiating into a subject's object file on itself (Peacocke, 2012b, pp. 79-80). For example, I might experience myself as being located relative to a particular environmental feature, and also experience myself as engaged in a bodily action in which I move forward. I see that I am to the left of the pond and have action-awareness of moving forward (Peacocke, 2012b, 79). These experiences are not isolated from one another, but it is not just the case that they are both part of my stream of consciousness. The two pieces of information are about me, and there is an integration of that information about me such that I experience myself as to the left of the pond and moving forward.

Peacocke explains this process of integration as a sub-personal mechanism, operating on representations of properties that we experience ourselves as instantiating. An integrating apparatus takes the relevant information from our sensory and perceptual systems and collates that information in the self-file. This process must be sub-personal: we do not engage in an activity of unifying the information we have on ourselves together, although of course that does not mean that we cannot engage in a similar process at the conscious level. The integration process that is a necessary condition on a subject's having self-representations operates below the conscious level:

"The seemings of perception, memory, emotion, and the rest are things that, at the conscious level, just happen to the thinker. Subject-reflexive perceptual seemings whose content concerns *de jure* the thinker himself are just a special case of this more general fact." (Peacocke, 2012b, p. 83)

The idea seems to be that it doesn't make sense to talk of self-representation at the non-conceptual level without it being the case that the relevant predicative contents are integrated into a self-file. What it is to be a self-representer, then, is to experience the world such that information about oneself is integrated together. One does not represent the spatial relation of a pond relative to oneself and one's current movement as one runs straight ahead in isolation from each other. Self-representation brings with it the integration of information about oneself.

6.6 Peacocke and Evans on first-person thought

As we saw in Chapter Five, Evans (1982) claims that we must employ an objective concept in first-person thought if we are to refer to ourselves when we entertain those thoughts. An objective self-concept will be a concept of oneself as a material being located in time and space. Evans requires that a subject integrates information from her experiences and engages in theorizing about where she is located. When a subject has a theory of perception she will have the capacity to locate herself in space, and thus will meet Evans's strong conditions on successful thought. But Evans's account is overly sophisticated, requiring subjects to engage in intellectual work if they are to have a conception of themselves as ordinary objects in the world. And this intellectual work must be carried out simply so that a subject is able to refer to herself. One outcome of this is that reference failure is a real and recognised possibility on Evans's account. I will look at this shortly.

A further outcome is that a subject must conceive of herself as a spatio-temporally continuous object if she is to refer to herself. She must, to some extent, keep track of herself perceptually as she moves around the world. Recall Evans's description of the kind of reasoning that will be required for a subject to have the capacity to locate herself in the objective world:

> "'I perceive such-and-such, such-and-such holds at *p*; so (probably) I am at *p*'; 'I perceive such-and-such, I am at *p*, so such-and-such holds at *p*'; 'I am at *p*, such-and-such does not hold at *p*, so I can't really be perceiving such-and-such, even though it appears to me that I am'; 'I was at *p* a moment ago, so I can only have got as far as *p*', so I should expect to perceive such-and-such'. These arguments exploit principles connecting the subject's position, the course of his perceptions, and the speed and continuity of his movement through space; and the child must learn to trip round and round those principles, so that he comes to think effortlessly in these ways." (Evans, 1982, p. 223)

Quite literally, Evans requires that a subject is able to keep track of herself as she moves around the world—updating her cognitive map of her surroundings and her relation to those

surroundings—if she is to be able to entertain first-person thoughts at all. She might employ a non-fundamental identification of herself in first-person thought, of course, but this still must be a conception of herself as a spatio-temporally located, continuous object: an object that could be located in space at a time.

The spatial (and temporal) reasoning that a subject must be able to engage in if she is to entertain first-person thoughts also allows her to think about the world as a unified space. As Evans says:

"The very idea of a perceivable, objective spatial world brings with it the idea of the subject as being *in* the world, with the course of his perceptions due to his changing position in the world and to the more or less stable way the world is. The idea that there is an objective world and the idea that the subject is somewhere cannot be separated, and where he is is given by what he can perceive." (Evans, 1982, p. 222)

According to Evans we can integrate information about the world into a unified whole by means of relating the things we experience in the world to ourselves. A door can be seen as straight ahead and a chair seen as to the left because the spatial locations of these two objects are related to the subject's own position. We can integrate visual information about a chair as being to the left with auditory information about a telephone ringing, again, by relating these objects to ourselves. What is required for this integration, suggests Evans, is for us to conceive of ourselves as a part of the objective order: one needs a conception of oneself as a material object in order to carry out this synthesis.

However, both the idea that one must explicitly engage in spatial reasoning of the kind described above, and the idea that one must conceive of oneself as a material object in order to think of the world as a unified whole both seem implausible. For Evans one must keep track of oneself as a material object—an object located in space and time—if one is to both have the capacity to entertain first-person thoughts and to think of the objective world. By contrast, according to Peacocke (2012b) we do not need to keep track of ourselves, nor must we think of ourselves as a material object, if we are to have either of these abilities. There are conditions on a subject being able to refer to herself in the special first person way, both at the conceptual level and the non-conceptual level, but these conditions are much weaker than Evans's conditions on successful thought. Moreover, for Peacocke the experience of the world as spatially unified, which the integration of information about oneself affords, does not depend on tracking oneself perceptually. It requires nothing more than that the content of experience have a non-conceptual *de se* or first-person content as a component.

According to Peacocke we explain a subject's experience of how things are with her now, and her experience of the world as spatially unified, by means of an integrative

apparatus that functions below the level of consciousness, holding together information about the subject in a self-file. The self-file stores mental representations about the subject, determining how the subject experiences herself as being both at the present time and over time (Peacocke, 2012b, pp. 79-80). Incorporation into the self-file will depend on a mental representation having first-person content as a component. The integration of information about the subject will therefore depend on the subject's enjoying states with first-person content. However, we should not take that integration to depend on a subject keeping track of herself, or thinking of herself as a particular material object. On the contrary, the nature of first-person content is such that information can be integrated without any identity conditions being employed. As Peacocke says in 'Explaining *De Se* Phenomena' (2012a):

"[…] to attribute *de se* contents in these basic action and perception cases is not at all to imply that the subject needs to keep track of the objects picked out by the *de se* component of the intentional content of the event. On the contrary, precisely in part because the first person is not a perception-based or a sensation-based demonstrative, but is rather an experience-independent indexical notion in its own right, there is no question of keeping track of what is thought about or represented as oneself. Keeping track of something is fundamentally a perceptually based capacity […]." (Peacocke, 2012a, p. 155)

For Peacocke we can explain both the subject's capacity for self-reference (at the non-conceptual level), and her experience of the world as spatially unified and herself as located within that world, without ascribing to the subject an objective conception of herself, nor the capacity to keep track of herself through perceptual reasoning. All we need to do is postulate the existence of an integrative mechanism, operating sub-personally, that collates information about the subject in a self-file. It is by integrating information in a self-file that we can represent ourselves as instantiating the conjunction of many properties, as standing to the left of the chair and in front of the door, for example, or as being to the left of the chair and as moving forwards towards the door (Peacocke, 2012b, p. 79).

Furthermore, Peacocke only requires that suitable information be integrated in the subject-file where that information is forthcoming. It is not the case that any particular information or source of information is necessary for the subject to have self-representations. By contrast, Evans's account of conceptual first-person thought makes a subject's capacity to entertain first-person thought dependent on the subject having, at some point, perceptual experiences with spatial content. Peacocke says:

"One apparent attraction of this approach to primitive self-representation is that it gives priority neither to perception, nor to thought, nor to action, nor to sensation in an account of pimitive self-representation. A subject that has perception, but no action-awareness, can meet this condition. So can a Helen Keller. Provided the subject can enjoy states and events with subject-reflexivity, and the predicative content of those states is transmitted to an

object-file on the subject, then the conditions for having this primitive nonconceptual representation of the subject are fulfilled." (Peacocke, 2012b, p. 85)

Certainly spatial information received through the sensory systems will typically be integrated into the subject's object-file on itself, but this does not mean that any priority is given to information received in a particular way.

The requirement that a subject think of herself as a material object located at a particular position in space and time has the consequence for Evans that it is possible for a subject to fail to refer to herself. Evans notes:

"I do not see, then, that it is absurd to suppose that there might be a subject of thought who is not in a position to identify himself, and whose attempts at self-identification fail to net any object at all." (Evans, 1982, p. 253)

Peacocke's account of first-person thought, by contrast, has the benefit of all accounts of the self-concept according to which the reference of the concept is governed by the self-reference rule: it has guaranteed successful reference.[41] Peacocke gives the following break down of why this is the case:

"The fundamental reference rule for *I*,
(FRR *I*) A use of *I* in a thinking refers to its author
together with the principle
(15) Any thinking has an author
jointly imply that
(16) any use of I in a thinking refers.
That is, they jointly imply that there are no uses of *I* in a thinking that fail to refer." (Peacocke, 2010, p. 103)

For Evans it is a real possibility that a subject might think that she has referred to herself in thought, but in fact she has failed to do so because she does not have or has not employed a concept of herself as a material object. In Section 5.8 of Chapter Five we looked at the possibility of a brain in a vat entertaining first-person thoughts. A brain that has been envatted for the whole time that it has been functioning and that has only received illusory perceptual information about its environment will not be able to entertain first-person thoughts according to Evans, even though it will have the illusion of being able to locate itself in space and time by means of engaging in reasoning based on its illusory perceptual experiences (Peacocke, 2010, pp. 105-6). Peacocke argues that it is a mistake to think this envatted brain cannot refer to itself first-personally.

"This verdict is not intuitive. While our subject's attempted perceptual demonstratives *this table*, *that mountain*, and the like will fail to refer if he is permanently envatted, it seems highly plausible that in such thoughts as *I am suffering more from this pain than from that*

[41] See Section 3.4 of Chapter Three.

one, and *If these experiences are all illusory, I wonder if there is some way I can change my situation?*, the use of the concept *I* refers to the envatted subject doing the thinking." (Peacocke, 2010, p. 106)

It is then, he claims, an attraction of his view over Evans's, that it has the consequence that first-person thought cannot fail of reference. As an account of first-person thought, it provides us with a view of the self-concept as a concept of a subject and thinker. We are able to have an integrated experience of ourselves at a time and over time without drawing on any personal-level capacities other than the capacity to have non-conceptual self-representations because the integration of information about oneself takes place at the sub-personal level. All that is required for this integration to take place is for the relevant mental representations to have first-person content. This first-person content refers *de jure* to the subject of the experience, so one need not represent or refer to oneself as a material object for the integration to occur.

6.7 The metaphysics of mind and a satisfactory account of first-person thought

In Chapter One I argued that a satisfactory account of first-person thought must explain the self-consciousness of such thoughts. In my thesis I have argued that it is difficult to provide an account of first-person thought that meets this condition. However, in this chapter I have presented Peacocke's (2010) explanation of fully self-conscious thought, an account which we can combine, as he does, with the view that the first-person concept is individuated by the self-reference rule, to provide an account of first-person thought that accommodates both the self-consciousness and also the guaranteed referential success of first-person thought.[42] According to Peacocke we can explain the self-consciousness of first-person thought by ascribing to the subject a tacit grasp of the self-reference rule along with facts "common to any ordinary case of thinking" (Peacocke, 2010, p. 81). These facts are that a subject has action-awareness with first-person content when she performs mental actions such as judging and that she can take the content of that awareness at face value, making a judgement with the same content.

We can, then, take Peacocke's account to depend on three significant claims. The first is the claim that many mental events can be classed as types of action. The second is the claim that we have an awareness of what we are doing when we engage in these mental actions and

[42] See Chapter Three and Section 6.6.

that this awareness has first-person content. The third is the claim that we can take the content of this awareness at face value and endorse the content in judgement.

Underlying Peacocke's account of fully self-conscious thought is a commitment to the idea that we are always able to self-ascribe judging to ourselves when we engage in an act of judging. He explains this capacity to self-ascribe judgements in terms of action-awareness with first-person content and the capacity to take the content of this action-awareness at face value. What Peacocke rejects, then, is Lichtenberg's (1990) response to the *cogito*. Lichtenberg famously suggests that Descartes is not justified to move from awareness of a conscious thinking to the judgement *I am thinking*. Rather, Descartes should only judge *there is thinking* or *it thinks* (Lichtenberg, 1990, p. 168). Peacocke must avoid this view of what it is that we are able to judge if he is to provide the account he does of one's knowledge that one is referring to oneself when one entertains a first-person thought. He must do so because a subject who is justified only in judging *there is judging I am tall* will not be able to infer that the token of the self-concept employed in her thought *I am tall* refers to herself.

It is Lichtenberg's view of what one is able to judge about one's judging that forms the basis for Anscombe's (1994) concern over the adequacy of a third-personal statement of the self-reference rule and O'Brien's (2007) objection of circularity. Both challenges depend on its being a real possibility that a subject is only able to judge *there is judging I am tall* when she makes the judgement *I am tall*. Both Anscombe's and O'Brien's objections are based on the assumption that an account of first-person thought must explain how a subject knows that she is referring to herself when she uses the self-concept. As we saw in Sections 6.2 and 6.4 this assumption is itself based on the claim that it is possible that a subject could entertain a thought in which a concept is employed whose reference is determined by the self-reference rule and yet the subject will not know that she is referring to herself. For Anscombe or O'Brien to have an objection to Peacocke's account, it must be the case that it is possible that a subject could make a judgement, *P*, yet be justified in judging only *there is judging that P*, rather than *I am judging that P*. If an ordinary subject is always in a position to make the latter judgement, as Peacocke suggests, then for any judgement whatsoever, including judgements in which the first-person concept is employed, the subject will be in a position to know that she is referring to herself.

In the remainder of this chapter I will examine the alternative view of the mind that would sustain the idea that a subject who judges *I am tall* will be entitled to judge only that *there is judging I am tall*, or more simply that a subject who judges that *P* will be entitled to judge only that *there is judging that P*. That is, I will determine what kind of view of the

mind Anscombe and O'Brien must be committed to if their objections are to carry through against Peacocke.

It is a requirement of Peacocke's account that mental events can be classified as mental actions, so that we can hold what he calls the principal hypothesis (Peacocke, 2010, pp. 253-254) and extend his account of how we come to know about our bodily actions—action-awareness—to an account of how we come to know about our judgements. This is required on Peacocke's account because a subject who knows the self-reference rule knows that a token of the self-concept refers to the *producer* of the thought in which it is employed. It will not be enough for a subject to have mere introspective awareness of her first-person judgement if she is to be in a position to know that she is referring to herself when she entertains a first-person thought.[43]

One way in which one might hold an alternative view of the mind to the view that Peacocke is committed to for his account of fully self-conscious thought to go through would be to take *all* mental events to be passive. In particular, one would have to hold that judgings, decidings and the like are not mental actions. The claim that judgings and other mental events are mental actions, has become increasingly popular (see, in particular, the essays in O'Brien & Soteriou, 2009). For example, O'Brien (2007) says:

"Much of our mental lives are active. When we think, judge, question, imagine, suppose, decide, deny, reflect, and so on, we are *prima facie* involved in mental actions or activities." (O'Brien, 2007, p. 75)

The alternative view of the mind will deny that these kinds of mental events should be classified as actions. It is a view of the mind that Galen Strawson (2003) promotes. Strawson does not deny that there are such things as mental events and that the mind is active. What he denies is that reasoning and judging should be classified as intentional actions (Strawson, 2003, p. 227):

[43] In fact, Peacocke gives an account of how a subject *can* know that a self-concept that occurs in a passive first-person thought refers to herself also (Peacocke, 2010, pp. 90-92). He accepts that not all thinking is active. Although we are in fact the producers of all our thoughts, some thoughts simply occur to us, rather than being the products of an exercise of mental agency. Yet, one can know that a token of the self-concept that occurs in a passive thought refers to oneself (Peacocke, 2010, p. 90). The reason for this is that, although one has no knowledge of performing an action, one can typically rely on its being the case that one has produced those passive thoughts:
"There are, then, in ordinary cases two sources of knowledge that your uses of *I* in your judgement refers to yourself. One involves knowledge that you are the agent of your judgement. The other, applicable in normal cases to all your thinkings, passive and active, involves knowledge that you are the person thinking them (no one else is the agent producing them). […] The explanation of the entitlement a normal thinker has for judging that in passive first-person thinkings he is referring to himself involves his right to rely in normal circumstances on his being the producer of his passive thinkings." (Peacocke, 2010, p. 91)

"Obviously thought involves—is—mental *activity*, but activity, whether mental, chemical or volcanic, does not always involve action. And if we consider things plainly, we find, I think, that most of our thoughts—our thought-contents—*just happen*." (Strawson, 2003, p. 228)

Strawson does not in fact deny that there is such a thing as mental action, but he denies that many of the mental events that we might think of as mental action, including judging and reasoning, should really be classed as such. Mental actions consist, he says, in something like concentrating one's mind on a particular problem or maintaining one's attention (Strawson, 2003, pp. 231-232):

"No doubt there are other such preparatory, ground-setting, tuning, retuning, shepherding, active moves or intentional initiations. But action, in thinking, really goes no further than this. The rest is waiting, seeing if anything happens, waiting for content to come to mind, for the 'natural causality of reason' to operate in one. This operation is indeed spontaneous, but in the sense of 'involuntary, not due to conscious volition'. There is I believe no action at all in reasoning and judging considered independently of the preparatory, catalytic phenomena just mentioned, considered in respect of their being a matter of specific content-production or of inferential moves between particular contents." (Strawson, 2003, p. 232)

The passive view of judgement and judging that Strawson endorses would have the result that a subject would not be in a position to know that she is a self-referrer when a first-person judgement occurs to her. If judgings and other mental events are not classed as mental actions, then we would not be able to argue that we come to know about those judgements through action-awareness. A subject's awareness of her judgements will be limited to introspective awareness, the kind of awareness she has of all her other passive mental states. This awareness might itself have first-person content, just as action-awareness is claimed to have first-person content. However, it will not have the content *I am judging I am tall*, say. The awareness will not have content to the effect that the subject is engaged in performing a mental action. She will not be aware of herself as being the producer of the judgement. But, as Peacocke himself says:

"It is not sufficient for reaching that knowledge [that one is a self-referrer] merely that a thinker have an awareness from the inside that a thinking is occurring, not even a thinking with a first-person content." (Peacocke, 2010, p. 89)

It is possible that a subject can gain this knowledge about her passive thoughts simply by relying on her background knowledge that she is the producer of her mental states, without any awareness that she is the producer of that passive state (Peacocke, 2010, 90-91). However, if all the subject's mental states are passive and she has no action-awareness at all, it is not clear how the subject will have the conception of herself as the producer of any of those states. The active route to self-knowledge will certainly not be available to her, but

neither will the passive route since she will not know that she is the producer of any mental states at all.

So, we can see that one alternative to Peacocke's view of the mind will involve denying that judging, reasoning and so on are types of mental action. The broadest alternative to Peacocke's view of the mind with respect to his account of fully self-conscious thought would, then, deny that any mental events are to be classed as actions. As such, there could be no action-awareness of producing any thoughts whatsoever, and no possibility of taking the content of action-awareness at face value.

This does not seem be the view of the mind embraced by Anscombe or O'Brien. Both object to Peacocke's account on the strength of it being possible for a subject to be in a situation in which she is not able to entertain the judgement *I am judging I am tall*, say. They do not claim that it is *never* the case that a subject can come to judge *I am judging I am tall* in the way that Peacocke elaborates. It is unlikely, then, that they will deny that any mental events are actions because a consequence of this view will be that a subject never has action-awareness with first-person content, and hence can never take the content of action-awareness at face value.[44]

A less extreme contrast with Peacocke's view of the mind classifies mental events such as the making of judgements as mental actions, but denies that we have action-awareness with first-person content. The alternative view of the mind might hold that we have no action-awareness of what it is that we are doing when we perform mental actions, or it might postulate the existence of action-awareness but deny that this form of awareness has first-person content.

According to the first option, the only awareness we have of our mental states and actions is introspective awareness. We come to know about our beliefs and desires through introspection, and we also come to know about our mental actions through introspection. On this view of the mind a subject who introspects her judgement will have knowledge of that judgement. However, she will not be aware that the judgement has been produced by herself. Obviously without action-awareness, she will not be able to take the content of her action-awareness at face value by judging *I am judging I am tall*.

However, the phenomenon of thought insertion might seem to tell against this view of our knowledge of our mental actions. In particular, it seems to illustrate the existence of

[44] Moreover, O'Brien expressly classes a variety of mental events as mental actions as the quotation cited above illustrates: "Much of our mental lives are active. When we think, judge, question, imagine, suppose, decide, deny, reflect, and so on, we are *prima facie* involved in mental actions or activities." (2007, p. 75).

action-awareness or a sense of agency over and above the introspective awareness that we have of our mental states. John Campbell (1999, 2002) argues that we should explain the phenomenon of thought insertion in terms of deficiencies in a sense of agency or the sense of having produced a particular thought, which we ordinarily experience when we perform a mental action. Campbell describes the experience of thought insertion in the following way:

> "The content of the experience seems to be exactly that token thoughts are being generated by some other person, and, perhaps with malice, inserted into the mind of the patient, so that the patient has direct introspective knowledge of a token thought which was generated by someone else." (Campbell, 2002, p. 36)

Thought insertion, then, seems to motivate a view of our knowledge of our minds that requires more than mere introspective awareness. If Campbell's description is accurate, a schizophrenic subject who experiences thought insertion is introspectively aware of thoughts in much the same way that we ordinarily come to know about our thoughts. What she lacks is the sense of production of her thoughts, and it is this absence of the sense of agency that accounts for her experience that those thoughts have been inserted into her mind by another agent.

Campbell bases his view on the work of Christopher Frith (1992), amongst others, who gives a comparator model-based account of thought insertion. On Frith's account we explain thought insertion using the same model that we use to explain the sense of agency that we have over our own bodily actions. We have a sense of agency when there is a match between the forward model—the predicted outcome of a particular action—and the sensory feedback about the state of oneself and the world once an action has been attempted. So too, a match between the thought that a subject introspects and the content of the efferent copy sent to the comparator generates a sense of agency over that thought (Campbell, 1999, p. 617). Thought insertion is explained, along with other schizophrenic symptoms, as a disturbance to the sense of agency generated by problems with "central monitoring" (Campbell, 1999, p. 617).

Peacocke also explains the phenomenon of thought insertion in terms of a subject's awareness of producing her judgements and other active thoughts.[45] Specifically, he takes the classification of judgements as mental actions, and the postulation of action-awareness as the means by which we know what we are doing to provide a partial explanation of the phenomenon of thought insertion. As Peacocke points out, it would be wrong to assume that schizophrenic patients take the first-person concept in their apparently inserted thoughts,

[45] I do not mean to suggest that there are no differences between Campbell's (1999, 2002) account of thought insertion and Peacocke's (2010) account. I only intend to draw attention to the similarities between their views, in particular to the idea that we can explain the phenomenon of thought insertion in terms of the subject lacking awareness of having produced a mental state.

where it occurs, to refer to themselves (2010, p. 89). A schizophrenic subject does not know that she is thinking about herself when she entertains a first-person thought that is experienced or taken by that subject to be inserted. Peacocke explains the phenomenon in terms of the absence of action-awareness (2010, p. 276). The schizophrenic subject has some kind of awareness of the thought, but without the accompanying action-awareness that enables her to judge that *I am judging that I am F*, say. Lacking the basis for making that first-person judgement about what she is doing, she is not in a position to take a token of the self-concept to refer to herself. Her tacit knowledge of the self-reference rule will involve the knowledge that a token of the self-concept refers to its producer. In the absence of action-awareness the schizophrenic subject entertaining a first-person thought will not take herself to be the producer of that particular token of the self-concept. More generally, without action-awareness she will not take herself to be the producer of the judgements for which she lacks action-awareness, and she will come to report that some of those judgements have not been produced by herself.

Even allowing, however, that there is such a thing as action-awareness, a view that still differs from Peacocke's will claim that our action-awareness of what we are doing does not have first-person content. Peacocke's analysis of the content of action-awareness is as follows:

"The content of the action-awareness is both first-personal and present-tensed. The content is of the form *I am doing such and such now*. When you take such an awareness at face value, and judge *I am doing such-and-such now*, your judgement is identification-free in a familiar sense. It is not the case that you are making this judgement only because, for some mode of presentation *m* other than the first person, you judge that *m* is doing such-and-such now, and you also accept that you are *m*." (Peacocke, 2010, p. 248)

But an alternative view of our knowledge of our actions might hold that action-awareness does not in fact have first-person content. If it does not, then even if the content of action-awareness is taken at face value, that will not of itself entitle the subject to self-ascribe, say, judging *I am tall*, as is required for a subject to know that she is thinking about herself. On this view our awareness of our mental events will have the content *there is judging I am tall*, perhaps, or *x is judging I am tall*, where *x* is a mode of presentation other than the first-person. We are, on this view, aware of mental actions as actions, but we are not aware of ourselves as being the producers of those actions.

Peacocke raises the identification-freedom of our first-person thoughts about what we are doing as evidence in favour of the claim that action-awareness has first-person content. It is not the case, he says, that one's first-person judgement *I am judging I am tall* is based on

the identification of oneself with the thing one is aware of as carrying out that particular mental action, identified by some mode of presentation other than the first-person.

Yet, without denying that first-person thoughts about what one is doing are identification-free, we can still query Peacocke's claim that this tells in favour of the idea that action-awareness has first-person content. We can account for the identification-freedom of first-person thoughts in two ways. Some identification-free thoughts are ones in which the use of the self-concept is representation dependent in the manner suggested by Peacocke that I described in Section 6.4, but a further set of identification-free thoughts, he says, are ones in which the use of the self-concept is representation *independent* (Peacocke, 1999, pp. 266-272).

If we accept that the thought *I am judging I am tall* involves a representationally dependent use of the self-concept, then we will be obliged to say that the content of action-awareness is first-personal. However, there are other thoughts that can be identification-free. Peacocke himself distinguishes representationally dependent uses of the self-concept from representationally independent uses (1999, pp. 266-272). A representationally independent use of the first-person is a use that is not made by taking the representational content of an experiential state 'at face value'. The self-concept is used even though there is no corresponding first-person content in the content of the state that grounds the first-person thought in which the self-concept is employed. According to Peacocke conscious states can be self-ascribed without the state having any first-person content. One can self-ascribe seeing a phone on a table simply because one sees a phone on a table, without the content of one's experience giving one reason to make that self-ascription (Peacocke, 1999, p. 266). What is more, a thought in which a representation independent use of the self-concept is made is identification-free (Peacocke, 1999, p. 270).

The notion of representationally independent uses of the self-concept gives us the idea of a non-content-based account of self-ascription (O'Brien, 2007, p. 108). What is more, these first-person thoughts are identification-free. I do not entertain the thought *I see the phone on the table* by identifying myself with the individual I am aware of as seeing the phone on the table, at least if I entertain that thought on the basis of having a visual perception of a phone on a table. The identification-freedom of first-person thoughts about what it is that one is doing, then, might be explained not by action-awareness having first-person content, but by these thoughts being non-content-based or representation independent thoughts. The identification-freedom of our first-person thought about what we are doing does discount the possibility that our awareness of our mental actions has the content *x is judging I am tall*. It

does not, though, discount the possibility that our awareness of our mental actions has the content *there is judging I am tall*. It leaves open the possibility that our awareness of our mental actions is not first-personal: we are not aware of ourselves as the agents of these mental actions, but only aware of those mental actions as occurring.

This is the view of the mind that Anscombe and O'Brien must adopt in order to make their objections stick. On this view there is action-awareness, but the content of action-awareness is not first-personal. An account of how we are able to make self-ascriptions of what we are doing will be needed in order to explain how, ordinarily, we can make self-ascriptions of what we are doing. This account must accommodate the fact that those self-ascriptive judgements will be identification-free. The suggestion, though, is that the self-ascriptions we make on the basis of action-awareness are not simply the result of taking the content of action-awareness at face value. Some more work is required on the part of the subject, and that is why it will be possible, in some cases, for a subject to be in a position where she can judge only *there is judging I am tall*. In these cases, the subject is not able to make the move to the self-ascriptive judgement *I am judging I am tall*, yet she is able to take the content of her action-awareness at face value and judge *there is judging I am tall*.

6.8 Conclusion

Peacocke's account of fully self-conscious thought involves a commitment to the idea that one has action-awareness with content to the effect that it is oneself that is judging that *P*, for any judgement *P* that one makes. From this Peacocke argues that a subject will always be in a position to know that she is referring to herself when she entertains a first-person thought because the content of her action-awareness will be of the form *I am judging I am F*. In this section I have examined what kind of view of the mind we would be committed to if we were to object to Peacocke's account in the manner of Anscombe or O'Brien. We have seen that, for their objections to work, we must view our knowledge of what we are doing when we perform mental actions to be based on action-awareness that does not have first-person content.

More generally, it seems that our ability to give an account of fully self-conscious thought is going to depend on the view of the mind that we adopt. In my thesis I have examined a variety of views of first-person thought. We have seen that it is particularly difficult to explain satisfactorily how a subject knows that she is thinking about herself when she entertains a first-person thought. Peacocke (2010), however, offers an account of how a

thinker of a first-person thought can be in a position to know that she is referring to herself. It is a condition on a subject thinking a fully self-conscious thought that she has a grasp of the self-reference rule. It will also, though, be a requirement that a subject knows that she is the producer of a particular first-person thought. For the many active first-person thoughts she entertains, it is a condition on a subject knowing that she is thinking about herself that she is aware that she is judging, imagining, and so forth. It seems that our ability to provide an explanation of the self-consciousness of first-person thought, therefore, will depend on our willingness to hold a particular view of the metaphysics of the mind: that the mind's relation to itself is essentially first-personal. It is not enough for a subject simply to be aware of a first-person judgement, she must be aware of that first-person judgement as one of her own.

In my thesis I have sought to provide an account of first-person thought that accommodates both the guaranteed referential success and the self-consciousness of the thoughts that we entertain about ourselves using the self-concept. We have seen that, although we apparently ascribe psychological properties to ourselves on the basis of introspection alone we do not and cannot introspectively observe a subject of thought and experience. The elusiveness of the subject of thought from introspective observation appears to generate a problem: not only have we not got an explanation of first-person thought, but it also appears as though neither the self-concept nor the first-person pronoun is referential.

In response to the problem of first-person reference I examined a descriptive account of first-person thought, but found that the account can only be adapted to explain the self-consciousness of first-person thought in such a way as to render the account circular. I also examined whether the self-concept might be a concept of a material object since, even though we do not introspectively observe the subject of thought and experience, we do perceive our bodies both exteroceptively and interoceptively. However, the proponent of this objective view of the self-concept cannot make sense of questions about our embodiment expressed in the first-person.

I have argued, though, that we can provide an account of first-person thought that explains the guaranteed referential success and self-consciousness of those thoughts by making use of the self-reference rule. Reference that is determined by the self-reference rule will have guaranteed referential success. Moreover, we have seen that the self-reference rule account can accommodate the self-consciousness of first-person thought. In fact, a subject's

grasp of the self-reference rule has a role to play in the explanation of how a subject knows that she is thinking about herself when she entertains a first-person thought.

In this final chapter we have seen that the problem of accounting for the self-consciousness of first-person thought seems to arise from holding a passive conception of the mind. If we adhere to a view of the mind as passive, a view of mental states and events as simply occurring to a thinker, then we have a real problem in explaining how a subject can know that she is thinking about herself when she entertains a first-person thought. Hence we have a problem explaining why first-person thoughts have a particular cognitive significance. I suggested that Anscombe and O'Brien would not themselves adopt a view of the mind as passive if they want to suggest only that it is a *possibility* for a subject to entertain what seems to be a first-person thought yet for that subject to fail to know that she is thinking about herself. By contrast, Strawson (2003), who advocates a passive, Humean view of the mind, does not seem to be in a position to explain fully self-conscious thought.

What Peacocke provides is a way of explaining fully self-conscious thought by arguing for a different perspective on the mind. Accepting that the mind is active and that we are aware of ourselves as engaged in the production of mental states allows us to provide this explanation. Questions about whether any mental events should be classed as mental actions, and how we come to know about our mental states and events take us beyond the scope of this thesis. It is significant, nonetheless, to know that our ability to explain the cognitive significance of first-person thoughts—thoughts that seem to be an integral part of much of our adult cognitive lives—will depend on our being able to argue for the conclusion that the mind is active.

References

Allison, H. E. (1983). *Kant's transcendental idealism: An interpretation and defence*. New Haven; London: Yale University Press.

Almog, J., Perry, J., & Wettstein, H. K. (Eds.). (1989). *Themes from Kaplan*. New York ; Oxford : Oxford University Press.

Anscombe, G. E. M. (1994). The first person. In Q. Cassam (Ed.), *Self-knowledge* (pp. 140–159). Oxford: Oxford University Press. (Original work published 1975).

Armstrong, D. M. (1984). Consciousness and causality. In D. M. Armstrong & N. Malcolm (Eds.), *Consciousness and causality: A debate on the nature of mind*. Oxford: Basil Blackwell.

Atkins, K. (2005a). Commentary on Hume. In K. Atkins (Ed.), *The self and subjectivity* (pp. 33-36). Oxford: Wiley-Blackwell.

Atkins, K. (Ed.). (2005b). *The self and subjectivity*. Oxford: Wiley-Blackwell.

Bar-On, D. (2004). *Speaking my mind: Expression and self-knowledge*. Oxford: Oxford University Press.

Bermúdez, J. L. (1998). *The paradox of self-consciousness*. Cambridge, Mass.: MIT Press.

Bermúdez, J. L. (2003). The elusiveness thesis, immunity to error through misidentification, and privileged access. In B. Gertler (Ed.), *Privileged access: Philosophical accounts of self-knowledge* (pp. 213–231). Aldershot: Ashgate Publishing, Ltd.

Bermúdez, J. L. (2011). Truly understood, by Christopher Peacocke. *Mind*, *120*(480), 1276–1280.

Bermúdez, J. L., Marcel, A., & Eilan, N. (Eds.). (1995). *The body and the self*. Cambridge, Mass.: MIT Press.

Brewer, B. (1995). Bodily awareness and the self. In J. L. Bermúdez, A. Marcel, & N. Eilan (Eds.), *The body and the self* (pp. 291–309). Cambridge, Mass.: MIT Press.

Brook, A., & DeVidi, R. C. (Eds.). (2001). *Self-reference and self-awareness*. Amsterdam: John Benjamins Publishing.

Campbell, J. (1994). *Past, space, and self*. Cambridge, Mass.; London: MIT Press.

Campbell, J. (1999). Schizophrenia, the space of reasons, and thinking as a motor process. *The Monist*, *82*(4), 609–625.

Campbell, J. (2002). The ownership of thoughts. *Philosophy, Psychiatry, & Psychology*, *9*(1), 35–39.

Campbell, J. (Forthcoming). Lichtenberg and the cogito. *Proceedings of the Aristotelian Society*, *112*(3). Draft retrieved from http://www.aristoteliansociety.org.uk/pdf/Campbell.pdf

Cassam, Q. (Ed.). (1994). *Self-knowledge*. Oxford: Oxford University Press.

Castañeda, H.-N. (1966). 'He': a study in the logic of self-consciousness. *Ratio*, *8*, 130–157.

Castañeda, H.-N. (1969). On the phenomeno-logic of the I. *Proceedings of the 14th International Congress of Philosophy*, III: 260-266.

Chisholm, R. M. (1969). On the observability of the self. *Philosophy and Phenomenological Research*, *30*(1), 7–21.

Coliva, A. (2003). The first person: Error through misidentification, the split between speaker's and semantic reference, and the real guarantee. *The Journal of Philosophy*, *100*(8), 416–431.

Coliva, A. (Ed.). (2012). *The self and self-knowledge*. Oxford: Oxford University Press.

De Gaynesford, M. (2006). *I: The meaning of the first-person term*. Oxford: Oxford University Press.

Descartes, R. (1986). *Meditations on first philosophy: With selections from the objections and replies*. (J. Cottingham, Trans.). Cambridge: Cambridge University Press. (Original work published 1641).

DeVidi, R. (2001). Frege and the first person. In A. Brook & R. DeVidi (Eds.), *Self-reference and self-awareness* (pp. 31–48). Amsterdam: John Benjamins.

Evans, C. O. (1970). *The subject of consciousness*. London: George Allen & Unwin.

Evans, G. (1982). *The varieties of reference*. (J. McDowell, Ed.). Oxford: Clarendon Press.

Evans, G. (1985a). Understanding demonstratives. In his *The collected papers of Gareth Evans* (pp. 291–321). Oxford: Oxford University Press.

Evans, G. (1985b). *The collected papers of Gareth Evans*. Oxford: Oxford University Press.

Frege, G. (1993). On sense and reference. (M. Black, Trans.). In A. W. Moore (Ed.), *Meaning and reference* (pp. 23–45). Oxford: Oxford University Press. (Original work published 1892).

Frith, C. (1992). *The cognitive neuropsychology of schizophrenia*. Hillsdale, NJ: Erlbaum.

Garrett, B. (1998). *Personal identity and self-consciousness*. London: Routledge.

Gertler, B. (Ed.). (2003). *Privileged access: Philosophical accounts of self-knowledge*. Aldershot: Ashgate Publishing, Ltd.

Gertler, B. (2011). *Self-knowledge*. New York: Routledge.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Grush, R. (2007). Evans on identification-freedom. *Canadian Journal of Philosophy*, *37*(4), 605–617.

Hahn, L. E. (Ed.). (1998). *The philosophy of P.F. Strawson*. Chicago; Lasalle: Open Court.

Howell, R. J. (2006). Self-knowledge and self-reference. *Philosophy and Phenomenological Research*, *72*, 44–70.

Howell, R. J. (2010). Subjectivity and the elusiveness of the self. *Canadian Journal of Philosophy*, *40*(3), 459–483.

Hume, D. (2000). *A treatise of human nature: Being an attempt to introduce the experimental method of reasoning into moral subjects*. (D. F. Norton & M. J. Norton, Eds.). Oxford: Oxford University Press. (Original work published 1739-1740).

Kant, I. (1929). *Critique of pure reason*. (N. Kemp Smith, Trans.). London: Macmillan. (Original work published 1781/1787).

Kaplan, D. (1989a). Demonstratives. In J. Almog, J. Perry, & H. K. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). New York; Oxford: Oxford University Press.

Kaplan, David. (1989b). Afterthoughts. In J. Almog, J. Perry, & H. K. Wettstein (Eds.), *Themes from Kaplan* (pp. 564–614). New York; Oxford: Oxford University Press.

Kriegel, U. (2007). Self-consciousness. *Internet Encyclopedia of Philosophy*. Retrieved from http://www.iep.utm.edu/self-con/

Kripke, S. A. (1980). *Naming and necessity*. (Rev. and enlgd. ed.). Oxford: Blackwell.

Lewis, D. (1983a). Attitudes *de dicto* and *de se*. In his *Philosophical papers* (Vol. 1, pp. 133–156). New York: Oxford University Press, USA.

Lewis, D. (1983b). *Philosophical papers, volume 1*. New York: Oxford University Press, USA.

Lichtenberg, G. C. (1990). *Aphorisms*. (R.J. Hollingdale, Trans.). London; New York: Penguin Classics. (Original work published 1801).

Mach, E. (1897). *Contributions to the analysis of the sensations*. (C. M. Williams, Trans.). Chicago ; London: Open Court.

McDowell, J. (1998). Referring to oneself. In L. E. Hahn (Ed.), *The philosophy of P.F. Strawson* (pp. 129–145). Chicago; Lasalle: Open Court.

Millikan, R. G. (1990). The myth of the essential indexical. *Noûs*, *24*(5), 723–734.

Milne, A. A. (1926). *Winnie-the-Pooh*. London: Methuen.

Moore, A. W. (Ed.). (1993). *Meaning and reference*. Oxford: Oxford University Press.

Noonan, H. (Ed.). (1999). *Routledge philosophy guidebook to Hume on knowledge*. London: Routledge.

O'Brien, L. F. (1994). Anscombe and the self-reference rule. *Analysis*, *54*(4), 277–281.

O'Brien, L. (1995). Evans on self-identification. *Noûs*, *29*(2), 232–247.

O'Brien, L. (2007). *Self-knowing agents*. Oxford: Oxford University Press.

O'Brien, L., & Soteriou, M. (Eds.). (2009). *Mental actions*. Oxford; New York: Oxford University Press.

Peacocke, C. (1983). *Sense and content : Experience, thought, and their relations*. Oxford; New York: Oxford University Press.

Peacocke, C. (1999). *Being known*. Oxford; New York: Oxford University Press.

Peacocke, C. (2010). *Truly understood*. Oxford: Oxford University Press.

Peacocke, C. (2012a). Explaining *de se* phenomena. In S. Prosser & F. Recanati (Ed.), *Immunity to error through misidentification: New essays* (pp. 144–157). Cambridge: Cambridge University Press.

Peacocke, C. (2012b). Subjects and consciousness. In A. Coliva (Ed.), *The self and self-knowledge* (pp. 74–101). Oxford: Oxford University Press.

Perry, J. (1977). Frege on demonstratives. *Philosophical Review*, *86*(4), 474–497.

Perry, J. (1979). The problem of the essential indexical. *Noûs*, *13*(1), 3–21.

Perry, J. (1988). Symposium papers, comments, and an abstract: cognitive significance and new theories of reference. *Noûs*, *22*(1), 1–18.

Perry, J. (1998). Myself and 'I'. In M. Stamm (Ed.), *Philosophie in synthetischer absicht* (pp. 83–103). Stuttgart: Klett-Cotta.

Prosser, S., & Recanati, F. (Eds.). (2012). *Immunity to error through misidentification: New essays*. Cambridge: Cambridge University Press.

Ramsey, I. T. (1955). The systematic elusiveness of `I'. *The Philosophical Quarterly*, *5*(20), 193–204.

Roessler, J. (2009). Critical notice of Lucy O'Brien, Self-knowing agents. *Philosophical Books*, *50*(4), 227–234.

Russell, B. (1914a). On the nature of acquaintance, part I: preliminary description of experience. *The Monist*, *24*(1), 1–16.

Russell, B. (1914b). On the nature of acquaintance, part II: neutral monism. *The Monist*, *24*(2), 161–187.

Russell, B. (1914c). On the nature of acquaintance, part III: analysis of experience. *The Monist*, *24*(3), 435–453.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson's University Library.

Sartre, J.-P. (1960). *The transcendence of the ego: An existentialist theory of consciousness*. (F. Williams & R. Kirkpatrick, Trans.). New York: Hill and Wang. (Original work published 1937).

Schopenhauer, A. (1969). *The world as will and representation, volume 1*. (E. F. J. Payne, Trans.). New York: Dover Publications. (Original work published 1818).

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

Shoemaker, S. (1968). Self-reference and self-awareness. *The Journal of Philosophy*, *65*(19), 555–567.

Shoemaker, S. (1970). Persons and their pasts. *American Philosophical Quarterly*, *7*(4), 269–285.

Shoemaker, S. (1986). Introspection and the self. *Midwest Studies In Philosophy*, *10*(1), 101–120.

Shoemaker, S., & Swinburne, R. (1984). *Personal identity*. Oxford: Wiley-Blackwell.

Stamm, M. (Ed.). (1998). *Philosophie in synthetischer absicht*. Stuttgart: Klett-Cotta.

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

Strawson, G. (2003). Mental ballistics or the involuntariness of spontaneity. *Proceedings of the Aristotelian Society*, *103*(1), 227–256.

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

Strawson, P. F. (1994). The first person—and others. In Q. Cassam (Ed.), *Self-knowledge* (pp. 210–215). Oxford: Oxford University Press.

Wittgenstein, L. (1958). *Preliminary studies for the 'Philosophical Investigations' generally known as the blue and brown books*. Oxford: Basil Blackwell.

Wittgenstein, L. (2001). *Tractatus logico-philosophicus*. (D. F. Pears & McGuinness, Trans.). London; New York: Routledge. (Original work published 1921).