

The First Person and Self-Concern

Antroula Kafa

MPhilStud Philosophical Studies

UCL

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

Abstract

This is a thesis about the practical, affective, and normative significance of first person thought. I urge that theories of first person thought must pay closer attention to these dimensions of the phenomenon, and I seek to enrich our appreciation of these dimensions by proposing that first person thought is associated with a distinctive concern for self. My aim is to determine what account of first person thought would best explain its distinctive association with self-concern.

Chapter 1 introduces and motivates the project. Chapter 2 proposes a diagnostic category of ‘Reductionist Views’, according to which first person thought is adequately characterised by its fundamental reference rule. I argue that Reductionist Views do not contain the resources to vindicate the practical, affective, and normative significance of first person thought, and that they are independently implausible. In Chapter 3, I consider ‘Perceptual Views’ of first person thought. I argue that, although such views face important problems, they nevertheless provide some resources for explaining the cognitive significance of first person thought. Taking cue from Perceptual Views, in Chapter 4, I articulate the ‘Immediate Access View’, according to which first person thought is grounded in immediate self-knowledge and a nonlinguistic and nonconceptual form of self-awareness. I conclude by discussing some of the ways in which self-concern could be explicated along the lines of the Immediate Access View, and by identifying the general form of an account on which self-concern can be vindicated.

Impact Statement

In focusing on the practical, affective, and normative dimensions of first person thought, this thesis contributes to an appreciation of its relevance for our lives, our interpersonal relations, and our ethics. It also aims to further such appreciation, by establishing an association between first person thought and the important phenomenon of self-concern. Moreover, in arguing that our accounts of first person thought should take care to accommodate its evaluative components and its association with self-concern, the thesis attempts to orient the philosophical literature on the first person towards directions that matter.

Acknowledgments

My utmost gratitude goes to Dr. Fiona Leigh. Her understanding and support has been extraordinary. I would have given up on myself.

I am also grateful to Professor Lucy O' Brien for her generous assistance, and to Dr. Douglas Lavin for the endless and endlessly stimulating conversations.

Finally, I am grateful to my parents, for never failing to believe in me; and to my partner, Evita, who has shared my struggles tirelessly and valiantly, offering nothing but love.

Contents

| | |
|---|----|
| Contents | 6 |
| 1. First Person Thought, Significance, and Self-Concern..... | 8 |
| 1.1. Introduction | 8 |
| 1.2. First Person Thought..... | 11 |
| 1.2.1. Self-Conscious Reflexivity | 11 |
| 1.2.2. Freedom from Identification and Immunity to Error Through Misidentification | 13 |
| 1.3. Significance | 16 |
| 1.3.1. Practical Significance: The Messy Shopper..... | 16 |
| 1.3.2. Affective Significance: Hume’s Passions..... | 19 |
| 1.4. Self-Concern | 23 |
| 1.4.1. The Self-Concern Intuition..... | 23 |
| 1.4.2. The Project and Its Motivations | 26 |
| 2. The Reductionist View | 29 |
| 2.1. Introducing the View..... | 29 |
| 2.2. What About Self-Consciousness?..... | 33 |
| 2.2.1. The Insufficiency Worry and the Circularity Worry | 33 |
| 2.2.2. Nonconceptual Self-Awareness..... | 38 |
| 2.3. A Diagnosis..... | 42 |
| 2.4. Significance and Self-Concern | 46 |

| | |
|--|----|
| 3. The Perceptual View | 48 |
| 3.1. Introducing the View | 48 |
| 3.1.1. Evans' Account | 49 |
| 3.1.2. The Internal Perception Account | 53 |
| 3.2. Significance | 55 |
| 4. The Immediate Access View | 61 |
| 4.1. Introducing the View | 61 |
| 4.1.1. The Problem of Other Minds | 61 |
| 4.1.2. The View | 65 |
| 4.1.3. Some Notes | 69 |
| 4.2. Explaining the Self-Concern Intuition | 72 |
| 4.2.1. Immediate Self-Knowledge | 72 |
| 4.2.2. Pain and Pleasure | 78 |
| 4.2.3. Agency | 85 |
| 4.2.4. Results | 87 |
| Conclusion | 90 |
| Bibliography | 95 |

1. First Person Thought, Significance, and Self-Concern

1.1. Introduction

The topic of this thesis is first person thought, i.e. thought that is self-consciously about the thinker. First person thought is standardly expressed using the first person pronoun, as in propositions like ‘I feel pain’, ‘I should go out tonight’, or ‘I must fix my temper.’ As suggested by these examples, occasions for entertaining first person thoughts include introspection, deliberation and action-planning, and self-reflection (Guillot 2016: 138), though first personal thinking can generally be thought to comprise an ubiquitous feature of our conscious life.

Any use of ‘I’ (in thought or in speech) is an instance of first person reference, a phenomenon that is closely associated with first person thought. Although this thesis focuses on first person thought, given the intimate association between the two, it often also ventures into discussions of first person reference. Nevertheless, the two phenomena should not be confounded: while ‘first person reference’ describes the linguistic practice of referring to oneself using the first person pronoun, ‘first person thought’ describes all self-consciously self-referential forms of thought, linguistic or nonlinguistic, conceptual or nonconceptual.

First person thought – and reference – comprises a subject of intense philosophical debate, partly because it manifests certain peculiar properties which have proven difficult to explain. For instance, as already captured in our definition above, first person thought is guaranteed to be about the one who thinks it, and in such a way that the one who thinks it is guaranteed to be aware that it is about her. The various peculiarities of first person thought will be presented in the present chapter.

Several theories of first person thought have been advanced in philosophers' attempt to make sense of the phenomenon in light of its peculiarities, and a central objective of this thesis is to consider the extent to which they succeed. In Chapter 2, I consider an artificial category of 'Reductionist Views', according to which most of what there is to say about first person thought is explicable in terms of its fundamental reference rule. In Chapter 3, I discuss 'Perceptual Views', according to which first person thought can be explained in terms of a broadly perceptual relation that subjects stand in to themselves.

In addition to considering first person thought generally, my aim here is to focus on a particular aspect of its distinctiveness, namely its connection with behaviour and affect. It is commonly conceded that when agents represent themselves in first person thought, they operate under a distinctive, and distinctively practical, psychological apparatus. For instance, it has been observed that, when agents shift from non-first-personal to first-personal forms of thought, they often undergo significant psychological changes, which, in turn, influence their practical deliberation and action. Moreover, as it will be part of my goal to illustrate here, first person thought also exhibits a special connection with affect, as it is associated with distinctive emotional responses and phenomena.

My more specific objective is to home in on a particular aspect of the practical, affective, but also normative significance of first person thought: the association between first person thought and self-concern. The latter is illustrated through an observation to which I will refer as the 'Self-Concern

Intuition’, and which is most prominently voiced by Richard Wollheim in *The Thread of Life* (1984). Wollheim articulates the intuition by asking us to imagine a subject’s reaction to the thought ‘Someone (else) will tomorrow morning wake up blind’, and then to the thought ‘I will tomorrow morning wake up blind’. He then points out that the two thoughts are likely to elicit different reactions, with the subject finding the latter thought somehow more urgent or disconcerting than the former. According to Wollheim, this goes to show that representing events as bearing on our own well-being (i.e. first-personally), leads us to take a distinctive sort of concern in them, different from the concern we show for events which bear on the well-being of others.

A central argument of this thesis is that, if certain instances of first person thought are indeed associated with a distinctive concern for self, this is something that theories of first person thought ought to try to explain. Shining a light on the link between first person thought and self-concern serves to highlight the significance we attach to the former, by illuminating its practical importance and its bearing on ethical questions concerning the distinction between self and other, egoism and altruism.

Ultimately, I seek to answer the following question: what would first person thought *have to be*, in order for its practical and affective significance and its link with self-concern to make sense? My strategy in answering this question will be to consider the central candidates for an account of first person thought (the Reductionist View and the Perceptual View), and to reflect on whether, in addition to explaining first person thought more generally, each of them is able to accommodate its practical and affective significance and its association with self-concern. The aim is ultimately to arrive at an account of first person thought which is both promising in itself, and which makes sense of these dimensions of the phenomenon. In the present chapter, I introduce first person thought, focusing on its practical and affective significance and its association with self-concern. I also introduce the central tasks of the thesis.

1.2. First Person Thought

1.2.1. Self-Conscious Reflexivity

To begin with, first person thought is *reflexive*. First person thoughts refer to, or are about, the agent who thinks them; like a mirror, their reference reflects the identity of their user. The same can be said about first person reference. Assume that I, the author of this thesis, am experiencing some pain, and that I articulate this through the proposition ‘I am in pain’. Clearly, this is a proposition about me, and the ‘I’ therein refers to me. Now, if you, the reader, then say ‘I am in pain’, this second proposition, although identical to the one I had expressed, will be about you, and the ‘I’ therein will refer to you. Thus the reference of the first person pronoun shifts in accordance with, and tracks, the identity of its user. This, in turn, makes the first person pronoun an indexical, i.e. a term whose reference changes depending on the context in which it is employed.

What is more, in thinking first person thoughts, subjects always seem to be conscious, or aware, that they are thinking about themselves. Thus first person thought exhibits a special sort of reflexivity, which we may call ‘self-conscious reflexivity’. To better understand this feature, it is useful to consider cases of *non-self-conscious* reflexivity, i.e. cases whereby agents refer to themselves but are not aware that they are doing so. One such case is famously described by Ernst Mach in *The Analysis of Sensations* (1914). Mach describes entering a bus at night, seeing a man at the other end of it, and thinking something like, ‘That man is a shabby pedagogue’ (4n1). What Mach does not realise, however, is that he is looking into a mirror, and that he, himself, is the ‘shabby pedagogue’ he is thinking about. Now, in this situation, given that Mach is thinking about himself, his thought is reflexive, or self-referential. However, because he does not know that he is looking into a mirror, he is not *conscious* of the fact that he is thinking about himself. We may also say that, in this scenario, Mach is thinking about himself but not *as* himself. This makes his thought non-self-consciously reflexive.

The reflexivity that characterises Mach’s thought must be distinguished from the reflexivity that characterises first person thought.² As Marie Guillot puts it, when it comes to ‘I’-thoughts,

it is not just that the subject of the thought is identical with its object. That very fact is also part of what I grasp in having the thought ... someone who thinks “I am F” is somehow also aware of the reflexive relation, and of the identity between the two *relata*, the subject and object of the representation. (2016: 139)

As Guillot points out, it seems part of thinking first person thoughts that, in doing so, the subject is aware that she self-refers. As John Perry puts the point, unlike thoughts like Mach’s, first person thoughts are thoughts about oneself as oneself, rather than thoughts about the person one happens to be (Perry 2002: 209).

Hence theorists must explain, first, how first person thoughts come to refer to the one who thinks them, and, secondly, how they come to do so self-consciously. Moreover, whether a given account of first person thought vindicates its self-conscious reflexivity can serve as a measure of its success. If an account allows for instances in which first person thinking fails to refer to the thinker – that is, if it allows for first person thoughts that are about someone other than the thinker – it will not be a successful account of the phenomenon. The same goes for any account that explains the reflexivity of first person thought but not its self-consciousness.

When it comes to first person reference, a common strategy in accounting for reflexivity is to suggest that our uses of ‘I’ are governed by a semantic rule according to which the first person pronoun always refers to the one who uses it. Consider, for instance, Sydney Shoemaker:

[T]he rules governing the use of [‘I’] determine once and for all what its reference is to be on any given occasion of its use, namely, that its reference is

² The distinction has been articulated by a great number of theorists, including Frege (1918), Castañeda (1967, 2001), Anscombe (1981), Nozick (1981), and Evans (1982).

to the speaker, and leave no latitude to the speaker's intentions in the determination of its reference. (1968: 559)

Similarly, Jon Barwise and John Perry state that 'A reasonable thing to say about [the word *I*] is that, whenever it is used by a speaker of English, it stands for, or designates, that person' (1981: 670). P. F. Strawson also proposes that the linguistic practices governing our use of 'I' dictate that 'the first personal pronoun refers, on each occasion of its use, to whoever then uses it' (1994: 210). John Campbell sums up the intuition via the statement that 'Any token of 'I' refers to the one who produced it', which he calls the 'token-reflexive rule' for the reference of 'I' (1995: 102).

On the ensuing view, all we need to do to account for the reflexivity of first person reference is to take a look at its semantics, as it is simply part of correctly using the 'I' that, in doing so, one refers to oneself. But if the token-reflexive rule, as Campbell calls it, explains the reflexivity of 'I', does it also explain its self-consciousness? This question is addressed in Chapter 2, which considers a hypothetical category of 'Reductionist Views', according to which the token-reflexive rule plays a central or exclusive role in accounting, not just for first person reference, but also first person thought.

1.2.2. Freedom from Identification and Immunity to Error Through Misidentification

On a standard construal of reference, to refer is to – usually perceptually – pick out, or identify, a certain object, which comes to constitute the object of reference. On this picture, for instance, I refer to a chair in my room by picking out or identifying this chair as one of the objects in my environment. First person reference, however, does not seem to work in this way: it does not seem that, when I refer to myself (in thought or in speech) I do so by means of picking out myself, or identifying myself, as a physical object in my environment, or as one person among others. Indeed, it seems that even when my perceptual faculties are entirely obstructed, for instance when I am lying inside a sensory deprivation

tank, I can still successfully refer to myself (Anscombe 1981). This suggests that first person reference is, as philosophers put it, 'identification-free' (Bermudez 1998: 8, O'Brien 2007: 4-5). However, if first person reference does not work by identifying its object, as standard forms of reference do, how *does* it work? In what way does the 'I' latch onto its referent?

It is important to note that not *all* first-person judgments are identification-free: some judgments about ourselves do rest on identifying ourselves in the same way as we identify other objects of reference. To see this, let us revisit Mach's shabby pedagogue scenario. If, when Mach realises that he is staring into a mirror, he judges 'I am a shabby pedagogue', this judgment rests on Mach identifying himself as a certain kind of object – that is, as the man who looks like a shabby pedagogue. We can also put this another way. Mach's judgment that he is a shabby pedagogue could be seen as a result of his combining the judgment 'That man is a shabby pedagogue' with the judgment 'I am that man'. But the latter judgment clearly involves identification – it involves Mach identifying himself as that man in the mirror, the shabby pedagogue. In this case, then, and in other cases like it, the subject comes to judge 'I am *F*' by judging that *a* is *F*, and identifying *a* as herself (Peacocke 2014: 191).

Now contrast identification-based first-person judgments, such as Mach's, with judgments like 'I am in pain', or 'I see a tree'. Clearly, the latter are not reached by means of predicating something of someone, and then coming to identify that person as oneself. When I judge that I am in pain, I do not do so by thinking of *someone* who is in pain, and then identifying that person as myself; when I judge that I see a tree, I do not do so by means of identifying myself as *someone* who sees a tree. These judgments are identification-free: in making them, the subject knows that she is seeing a tree or feeling pain somehow directly.

Ludwig Wittgenstein (1958) maps the distinction between identification-free and identification-based judgments on to another distinction, between judgments that are immune to a certain sort of error and ones that are susceptible so it. Sydney Shoemaker, who elaborates on Wittgenstein's insight,

calls the relevant phenomenon ‘immunity to error through misidentification relative to the first-person pronoun’ (1968). Generally, ‘error through misidentification’ occurs when we think that something is true of someone, but we *misidentify* the individual of whom we take this to be true (Shoemaker 1968: 557). Suppose I come to believe, on the basis that the lights in my house are always left on, that my son is always leaving the lights on, when it is my wife. In this example, although I am correct in believing that someone is always leaving the lights on, I *misidentify* the person of whom this is true. I thus commit an error through misidentification.

The idea that certain uses of the ‘I’ are immune to error through misidentification is the idea that one cannot, in judging that certain things hold true of oneself, be mistaken in judging that it is *oneself* these things hold true of. Introspective judgments, like the judgment that one is in pain, are acknowledged as having this sort of immunity. The idea is that it is impossible, when I judge that I am in pain, to *misidentify* the person feeling the pain, mistakenly believing that it is me, when it is really someone else. Similarly, the judgment ‘I see a tree’ may be mistaken, but only because I may be wrong in believing that it is a *tree* that I see (I may instead be looking at an extremely realistic sculpture); not because I may be wrong in thinking that it is *me*, rather than someone else, who is seeing a tree (Shoemaker 1968: 557).

Though we have been talking about first person reference, immunity to error through misidentification also characterises first person thought and first-personal cognition in general. Consider a subject merely experiencing pain, without articulating her experience through a first person judgment. This experience is immune to error through misidentification because the subject cannot misidentify the subject of that experience – she cannot mistakenly think it is she herself who is feeling pain, when it is really someone else. Like first person judgments, first person thoughts and experiences are immune to error through misidentification, and this is another feature that our accounts of first person thought ought to explain.

1.3. Significance

1.3.1. Practical Significance: The Messy Shopper

There is a firm consensus in philosophy that first person thought exhibits a distinctive cognitive significance, consisting in its singular connection with practical reasoning and action. This idea is often advanced through consideration of the following example from John Perry's 'The Problem of the Essential Indexical' (1979):

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 ... I believed at the outset that the shopper with a torn sack was making a mess. And I was right. But I didn't believe that I was making a mess. That seems to be something I came to believe. And when I came to believe that, I stopped following the trail around the counter, and rearranged the torn sack in my cart. (3)

This thought-experiment exploits the distinction between self-conscious and non-self-conscious self-reference, or first personal and non-first-personal self-reference, previously discussed in relation to the shabby pedagogue case. When Perry thinks that the shopper with the torn sack is making a mess, he is thinking about himself but not *as* himself, i.e. in non-first-personal terms. When he comes to realise that he is the messy shopper, and thinks about himself knowing that it is himself he is thinking about, he enters the realm of first person thought. Crucially, once Perry enters the realm of first person thought, an immediate and notable change marks his behaviour: he stops following the trail of sugar and rearranges the torn sack in his cart. Therefore, his transition from a non-first-personal to a first personal form of self-representation has a direct impact on his behaviour. This impact constitutes the cognitive or practical significance of first person thought.

Importantly, just as ‘The shopper with the torn sack is making a mess’ fails to match the practical significance of ‘I am making a mess’, so would any other rendering of the latter in non-first-personal terms. To demonstrate this point, Perry considers replacing the ‘I’ in ‘I am making a mess’ with some other non-first-personal self-description, e.g. with ‘*John Perry* is making a mess’. He then points out that such a judgment would not influence his behaviour in the same way as ‘*I* am making a mess’, unless it were coupled with the judgment ‘*I* am John Perry’ (4-5). This suggests that resort to the first person is ultimately necessary in accounting for the change in Perry’s behaviour, and that attempts to articulate the relevant belief in non-first-personal terms while preserving its cognitive significance are bound to fail. As Perry puts it, any attempt to dispose of the ‘I’ by reducing it to a non-first-personal self-description would ‘destroy the explanation’ of its cognitive significance (5).

Thus the change in Perry’s behaviour is a unique function of his shifting to, and employing, a first personal form of thought. An important task facing theories of first person thought consists in explaining the practical significance of the phenomenon, as illustrated in cases like Perry’s. Why is it that when Perry shifts from a non-first-personal to a first-personal self-representation his behaviour is impacted in such a way? How can the change be characterised, or what is first person thought, such that it has this distinctive effect on behaviour and action?

Here, it must be mentioned that Perry’s own motivation in considering the messy shopper case is mainly to point out a problem with Fregean accounts of cognitive attitudes such as belief and thought. This problem is not directly relevant to my concerns here, so it will not be discussed in depth; nevertheless, a short perusal of it can help to set the groundwork for my discussion of the Perceptual View of first person thought in Chapter 3. Perry’s problem concerns the fact that, because of their distinct practical profiles, ‘The shopper with the torn sack is making a mess’ and ‘I am making a mess’ appear to be distinct beliefs. That is to say: given the change in Perry’s behaviour when he comes to believe that he himself is making a mess, it seems that what he comes to believe

in this case is different from what he formerly believed, i.e. that the shopper with the torn sack is making a mess. This is a problem for Fregean accounts of belief, according to which beliefs gain their meaning from, and are individuated by, their referents. For, since both these beliefs have the same referent (John Perry), they should, on the Fregean picture, be one and the same belief. So how is it that they behave so differently in relation to the agent's practical psychology?

Frege's philosophy of language does offer some resources for resolving this difficulty, namely by incorporating the concept of a 'sense' or 'mode of presentation' (*Art des Gegebenseins*) of a given object of reference (Torre 2016: 6). The basic idea is that, while the meaning of a thought is partly determined by what it refers to, there are also different ways of thinking about, or apprehending, or being presented with, the same referent (i.e. different 'senses'); and that the manner in which one is presented with the referent, in thinking about it, also contributes to the cognitive significance of the thought. So, in the case at hand, Frege could assert that 'The shopper with the torn sack is making a mess' and 'I am making a mess' may both have John Perry as their referent, however, they presuppose different ways of apprehending or being presented with that referent. That is, when Perry thinks 'The shopper with the torn sack is making a mess' and when he thinks 'I am making a mess' he may in both cases be thinking about himself, but he thinks of himself under a different 'sense' in each case. Since different senses have different sorts of cognitive significance attached to them, it is to be expected that when Perry shifts from one thought to the other, this is accompanied by a change in his behaviour.

This, however, raises a difficult question: how is Perry 'presented with himself' in each case? Most importantly for our purposes: how is Perry 'presented with himself' when he thinks about himself first-personally, and why does his being presented to himself in this way influence his practical deliberation in the way that it does? Famously, Frege has something to say about this as well. In 'Der Gedanke' (1918), he asserts that first person thought involves a subject being 'presented to himself in a particular and primitive way, in which he is presented to no-one else' (333). Of course, this remark only instigates

further confusion: what is this ‘particular’ and ‘primitive’ way in which we are presented with ourselves in first person thought? And, again, why does our knowing ourselves in this way have the cognitive significance that it does? These questions will be addressed in Chapter 3, where we consider Perceptual Views of first person thought (of which Frege’s view is one), and Chapter 4, where we present an account of first person thought which appeals to a primitive form of self-awareness.

1.3.2. Affective Significance: Hume’s Passions

To sum up: Perry’s remarks suggest that thinking about oneself through the first person has a distinctive impact on one’s action and deliberation – that there are distinctive attitudes and responses associated with, or springing from, first person thought. The Perry example, and other examples like it, are staples of the literature on the first person when it comes to illustrating the cognitive significance of first person thought. However, it can be doubted whether they firmly establish the point they intend to make. Before moving on to further elaborate the cognitive significance of first person thought, I will briefly address a few of these doubts.

First, one could offer an explanation that reduces the alleged cognitive significance of first person thought to its self-consciousness, i.e. to the fact that, in first person thought, subjects are aware that they are thinking about themselves. That is, one could argue that all Perry’s transition from a non-first-personal to a first-personal form of thought boils down to is a transition from not knowing that he is the messy shopper to knowing that he is, and that this is all we need to explain the change in his behaviour. On this picture, Perry stops and picks up the torn sack *because* he has come to know that he himself is making a mess, a fact of which he was not previously aware; and his acquiring this new knowledge is sufficient to explain the change in his behaviour. If it is a constitutive part of first person thought that the thinker is aware of her act of

self-reference, this alone provides us with enough resources to explain its alleged cognitive significance.

However, this response only seems to push the question a step back, for we may still ask: what is it that Perry comes to learn when he learns that he himself is the messy shopper? And why would his behaviour change upon acquiring this new piece of knowledge? After all, if self-consciousness is indeed a constitutive component of first person thought, to say that Perry's transition from a non-first-personal to a first-personal form of thought is a transition from a non-self-conscious to a self-conscious form of thought is not really to say much apart from reiterating a fact. We must still explain why this transition, even when couched in terms of the self-consciousness of first person thought, affects Perry's behaviour.

To this, one may retort that there is nothing special about the way Perry responds when he finds out that he himself is making a mess: this change in behaviour naturally follows from a change in belief, and such a change would take place no matter who the messy shopper turned out to be. If Perry had learnt that the shopper with the torn sack was his wife, he would then have reason to reprimand her about the mess she has made and help pick it up; if he'd learnt that a large man with a violent history has made the mess, he would have reason to quietly swerve away. In each case, Perry's behaviour would change in accordance with his change in belief, and there is nothing special about his picking up the mess when he finds out that he is responsible for it.

The latter seems like a potent objection, but we can push back by considering the relation between first person thought and a range of emotions like shame, embarrassment, guilt and pride, which, for ease of reference, and directly following David Hume, I will refer to as 'the passions'. Unlike other forms of thought, first person thought is associated with this domain of distinctive, and distinctively practical, emotions, which are exclusive to it, and which are bound up with its cognitive significance. In the case at hand, it is extremely likely that, when Perry finds out that he himself is the messy shopper, he will feel embarrassed, ashamed, self-conscious, or guilty about the trouble he

has caused. Crucially, these are not reactions Perry would exhibit upon learning that a violent stranger has caused the mayhem. In the latter case, he may feel angry or exasperated, but none of the passions will be relevant or even justified. Importantly, this suggests that the passions are only relevant within the realm of first person thought.

The connection between first person thought and the passions, particularly pride and shame, is extensively and forcefully expounded by David Hume in Book II of *A Treatise of Human Nature* (1738). First, Hume argues that the idea of self necessarily constitutes the *object* of pride and shame (which he refers to as ‘humility’), by being ‘that to which they direct their view, when excited’ (2.1.2.4).³ His insight here seems to be that the self, or something related to the self, always comprises that which I am proud or ashamed *of*, or *about*;⁴ on this picture, then, pride and shame are strictly first-personal. Hume views this as a cognitive fact about the passions, arguing that the idea of self is naturally embedded in their cognitive structure:

[Th]e peculiar object of pride and humility is determin’d by an original and natural instinct ... ‘tis absolutely impossible, from the primary constitution of the mind, that these passions shou’d ever look beyond self.’ (2.1.5.3)

Secondly, according to Hume, the idea of self is not only the object of the passions, but is also necessary for their arousal, thus constituting (part of) their cause. In Book II, Hume distinguishes between the *quality* and the *subject* of the causes of pride and humility; for example, according to Hume, if a beautiful house is the *cause* of one’s pride, the *subject* of the cause is the house, and the *quality* of the cause is beauty (2.1.2.6). Crucially, for our purposes, Hume thinks that the *subject* of the cause *must* in some way be related to the agent, in order to excite either pride or humility:

[T]is the beauty or deformity of our person, houses, equipage, or furniture, by which we are render’d either vain or humble. The same qualities, when

³ References to the *Treatise* are by Book, Part, Section and Paragraph number.

⁴ For a more detailed consideration of this claim, see Penelhum (1975).

transfer'd to subjects, which bear us no relation, influence not in the smallest degree either of these affections. (2.1.5.2)

So, if a beautiful house is the cause of my pride, the house must, in some way, be related to me: it might be my own, it might be my child's, or I might be the architect. In general, the subjects of the causes of pride comprise 'whatever objects are in the least ally'd or related to us' (2.1.2.5).

Thirdly, the passions' conjunction with the idea of self constitutes the crucial point of their distinction from a separate, non-first-personal class of emotions. At the beginning of Book II, Hume distinguishes between the 'direct' and 'indirect' passions; the former include 'desire, aversion, grief, joy, hope, fear, despair and security', and the latter 'pride, humility, ambition, vanity, love, hatred, envy, pity, malice, generosity' (2.1.1.4). Hume sums up the distinction between the two classes of emotions as follows:

By direct passions I understand such as arise immediately from good or evil, from pain or pleasure. By indirect such as proceed from the same principles, but by the conjunction of other qualities. (2.1.1.4)

Thus, according to Hume, what is common between the two classes of emotions is that they are both products of pain and pleasure, or good and evil.⁵ In the case of the direct passions, this means that an experience of pain or pleasure (or good and evil), like 'a fit of the gout', is all that is needed to bring them about (2.1.1.2). The indirect passions 'proceed from the same principles [as the direct ones]', which means that they are also, partly, caused by pain, pleasure, good, and evil. However, they are distinguished from the direct passions insofar as that their arousal also requires 'the conjunction of other qualities'; and, as it becomes clear through Hume's subsequent discussion, this refers to the idea of self. Therefore, Hume thinks that in cases where pain or pleasure are not accompanied by the

⁵ By 'good' and 'evil' Hume most likely means virtue and vice, which he defines as 'the good and bad qualities of our actions and manners' (2.1.5.2). The claim that virtue and vice constitute causes of the passions thus seems to be that one's perusal of such positive or negative qualities, either in oneself or in others, can cause one to feel joy, aversion, pride, humility, love, hatred, etc.

idea of self, the subject only experiences one of the direct passions (Penelhum 1976: 16). This goes to reinforce the point that the indirect passions of pride and shame are distinctively first-personal, since the idea of self is necessary for the arousal, and furthermore necessary in distinguishing them from the direct passions.⁶

Wrapping up, assuming Hume's key insights to be correct, we may say that emotions such as shame and pride belong exclusively to the domain of the first person. This entails that, in addition to its tight link with behaviour and action, which was illustrated by the messy shopper case, first person thought also possesses a distinctively *affective* significance, which consists in its association with a unique, and exclusively first personal, class of emotions. We may also put this by saying that the cognitive significance of first person thought possesses both a practical and an affective dimension. Importantly, the affectivity of first person thought also includes its connection with self-concern, which I will now go on to consider.

1.4. Self-Concern

1.4.1. The Self-Concern Intuition

Consider the following scenario from Richard Wollheim's *The Thread of Life* (1984):

So I am told something like the following: ... Someone whom I know will tomorrow morning wake up blind. Then I learn that this someone, the someone whom I know and to whom this will happen, is me. There is a characteristic ...

⁶ Hume's insights allow us to also observe that reactions of shame or embarrassment would also be relevant if the culprit for the mess was not Perry himself, but were in some other way related to Perry (it might be his wife or his dog). If Perry felt embarrassed upon learning that *his wife is* the messy shopper, this is because 'My wife is making a mess' is also a first person thought, despite its object not being John Perry.

way in which I shall respond to such a lesson. This response I call ‘the tremor’.
(237)

By contrast:

I am told that something good, something bad, will happen to someone whom I know. Then I learn who the someone is. It isn’t me this time, it’s a friend. The news is likely to affect me, and this response, which I shall call ‘the impact’, is to be distinguished, and non-arbitrarily, from the tremor. (237)

The insight is that thinking that a future evil will befall *me* is likely to produce a different response than thinking of the same evil befalling *another*. This insight constitutes the ‘Self-Concern Intuition’. Although Wollheim does not do much to describe the phenomenology of each response, we may assume, from the fact that he calls the one ‘the tremor’ and the other ‘the impact’, that he takes the former to be more forceful or acute than the latter. This could suggest that being confronted with the prospect of one’s own pain is more distressing than being presented with the prospect of another’s, i.e. that there is a difference *of degree* between the subject’s response in each case. However, if we take ‘the tremor’ and ‘the impact’ to each include a range of emotional responses, it is also possible that ‘the tremor’ incorporates responses like dread or terror which are not part of ‘the impact’, and thus that the difference between the two is not only one of degree. In that case, ‘the tremor’ could be said to be more *forceful* than ‘the impact’ insofar as the unique responses it involves, e.g. dread and terror, are especially acute by nature. Reversely, ‘the impact’ could involve responses like pity or sympathy, which do not readily apply to the first person case, and which are milder in character.

Characterising the exact difference between ‘the tremor’ and ‘the impact’, and the particular responses each involves, would be an interesting but extensive project. Here, it will be sufficient to restrict ourselves to pointing out that there is a *distinctiveness* in the way I am likely to respond to the thought of *my own* suffering, as opposed to another’s, which is not merely a difference of degree.

This is all that is needed for, and all that is asserted by, the Self-Concern Intuition.

The Self-Concern Intuition has also been articulated by Kieran Setiya in his recent paper ‘Selfish Reasons’ (2015):

[I]magine that I have forgotten my name. As I wake in the hospital ward, surrounded by other patients, I notice a medical chart that indicates a painful operation for Kieran Setiya later that day. I feel sorry for him and I hope he does not suffer too much. The doctor then walks by and informs me that I will have the same operation. I shudder in dread, much more distressed than I was before. Self-concern is addressed to my well-being from the first person perspective: to my well-being considered as mine, not just that of one among many. (445-446)

In this passage, Setiya proceeds along the same lines as Wollheim, by spelling out the Self-Concern Intuition through reflection on the distinction between ‘the impact’ and ‘the tremor’. When Kieran Setiya learns that Kieran Setiya will undergo a painful operation, he feels sorry for the poor guy, and hopes he does not suffer too much. He therefore manifests pity and sympathy – responses which, as suggested above, may be thought characteristic of ‘the impact’. By contrast, when he learns that he himself will have the painful operation, he shudders in dread, much more distressed than he was before. The latter reaction would certainly be called ‘tremorous’.⁷

Importantly, Setiya also addresses the distinctively first personal character of ‘the tremor’. As he points out, it is because he believes that the future evil will befall *him*, represented *as himself*, that he comes to ‘shudder in dread’ and become ‘much more distressed than [he] was before’. Setiya’s articulation of the example makes clear that we could not make sense of ‘the tremor’, and its distinction from ‘the impact’, without an appeal to the first person. When Setiya thinks that ‘Kieran Setiya will undergo a painful operation’, he does not exhibit any of the responses characteristic of ‘the tremor’, despite this thought being *about* him. Indeed, ‘the tremor’ only becomes possible, or relevant, when Setiya

⁷ As indicated by these remarks, Setiya construes ‘the tremor’ as involving responses which differ both in degree and in kind from those comprising ‘the impact’.

comes to represent the relevant event first-personally, in thinking ‘*I* will undergo a painful operation’. Hence ‘the tremor’ occurs uniquely as a response to the thought of my own future suffering, represented first-personally.

1.4.2. The Project and Its Motivations

The above remarks indicate that the practical and affective significance of first person thought also incorporates a distinctive concern for self, manifested in the cases articulated by Wollheim and Setiya, and captured in the Self-Concern Intuition.⁸ My interest here is in the explanation of the Self-Concern Intuition, and my aim is to determine what kind of account of first person thought we would have to adopt in order to vindicate the connection between first person thought and self-concern. The central question of this thesis, and the one it ultimately aims to answer, is the following: what would an account of first person thought have to look like if it were to do justice to, and explain, the Self-Concern Intuition?

The apparent connection between self-concern and first person thought bears on the latter’s significance for psychology, ethics, and intersubjectivity. Determining whether first person thought somehow incorporates a distinctive concern for self, and how, could illuminate some of our most fundamental attitudes towards ourselves and our lives.⁹ Moreover, a consideration of the connection between first person thinking and self-concern could be illuminating with regards to the possibilities and limitations of our interpersonal relations. Since we presumably cannot represent others in the same way as we represent ourselves, if (say) self-concern somehow flows from first person thought, where does concern for others come from? And how does it compare to self-concern, psychologically and ethically? For instance, what if first person thought is such that it compels the subject to care about herself *more* than she cares about

⁸ The association between first person thought and self-concern is also explored in Johnston (2010).

⁹ Like, for example, our fear of death (Johnston 2010).

others? Could certain naive convictions about psychological egoism then find their basis in the structure of first person thought?

As speculative as these questions may be, they point to philosophical directions that we would do well to explore. Indeed, the central question of this thesis, as stated above, implies that the demand to account for the Self-Concern Intuition should neither be optional, nor an afterthought – it should instead comprise a starting point in theorising first person thought. However, as I will discuss throughout this thesis, and particularly in Chapter 2, first person theorists have tended to underplay the affective and normative significance of first person thought, instead focusing many of their efforts on explaining features like its reflexivity. This thesis is a call to remind ourselves of the distinctively practical phenomenon that first person thought is, and to give it the corresponding philosophical treatment.

In the chapters that follow, I make a start at addressing the requirement set in this thesis by assessing our main accounts of first person thought with regards to whether, and how well, they can account for the Self-Concern Intuition. This approach will also serve a heuristic purpose, as it will eventually, and indirectly, help us determine what account of first person thought is best positioned to make sense of its connection with self-concern. Throughout my examinations, I will of course also be evaluating the general plausibility of the accounts under consideration, namely their success in explaining the other properties of first person thought as listed in the present chapter; i.e. its reflexivity, self-consciousness, immunity to error through misidentification, and general practicality and affectivity.

2. The Reductionist View

2.1. Introducing the View

In Chapter 1, I explained that an important desideratum on theories of first person thought is to account for the reflexivity of the phenomenon, i.e. for first person thoughts' guaranteed reference to the thinker. I also pointed out that reflexivity is an important feature of the linguistic first person, since the first person pronoun always refers to the one who employs it. In the case of first person reference, reflexivity is often accounted for by appeal to the 'token-reflexive rule', which dictates that 'Any token of 'I' refers to the one who produced it' (Campbell 1995: 102). Indeed, as the first person literature currently stands, the token-reflexive rule constitutes a highly favoured means of accounting for the reflexivity of first person reference. This is because, according to its champions, the rule provides a simple and highly economical explanation of the phenomenon: if the 'I' is semantically governed by a rule according to which it always refers to its user, and if competent use of 'I' is guided by this rule, it follows that each time a subject uses 'I' she will refer to herself.

The opinion is summed up by John Campbell as follows:

Only with reluctance should we take this step, which can seem so easy and indeed inescapable, of supposing that there is a use of the first person on which its reference is fixed not by the token-reflexive rule but in some other way. Many of the most distinctive phenomena involving the first person are straightforwardly explained by its being governed by the rule. Once we leave the

rule behind, problems swarm upon us. Most immediate and simple is the question of whether the alternative method of reference-fixing, whatever it is, is guaranteed to yield the same results as the token-reflexive rule in what references it finds for particular uses of the first person. If there is this guaranteed coincidence, then it is not apparent what advantage there can be in shifting to the new method. If the coincidence is not guaranteed, then we have opened up the possibility of someone using the first person to refer to someone other than himself. But this would not be recognizable as a use of the ordinary first person. (1995: 124)

Campbell's argument is the following: it is incumbent upon any account of first person reference to account for the reflexivity of the 'I'; an account of first person reference can do this either by accepting that the 'I' is governed by the token-reflexive rule, or by postulating some other way in which the 'I' secures its reference; since the token-reflexive rule constitutes the most effective, and secure, means of explaining the reference of 'I', there is no reason to look for an alternative explanation of its reference. The upshot is that searching for reference-fixing methods other than the token-reflexive rule is redundant, and precarious, given the availability and reliability of the token-reflexive rule.

Although this argument has a general application, Campbell has a particular target in advancing it. This is a picture on which the reference of 'I' is achieved via a perceptual relation to the self, invoked by Perceptual Views of first person reference. As we will see in the next chapter, a crucial problem with the method of reference-fixing employed by certain forms of the Perceptual View is that it fails to guarantee the reflexivity of the 'I' (and of first person thought in general) on every occasion of its use, because it allows for instances in which the first person can be used to refer to someone other than the user. As Campbell suggests in the passage, this failure of the perceptual method (or any method) entails that there is no reason to entertain it as a possible contender to views that employ the token-reflexive rule. On a more general – and more ambitious – note, Campbell also suggests that, even if a reference-fixing method is just as successful as the token-reflexive rule, there is hardly a reason to prefer it to the

latter. Presumably, this is because, Campbell thinks, the token-reflexive rule is the most secure and economical method of reference-fixing.¹⁰

Interestingly, when Campbell states that ‘Many of the most distinctive phenomena involving the first person are straightforwardly explained by its being governed by the rule’, he seems to be suggesting that the token-reflexive rule could also explain features of first person reference other than its reflexivity; presumably features like its self-consciousness and immunity to error through misidentification. The clear implication is that the rule need not simply comprise part of an account of first person reference, but that it might be all such an account needs. On a similar note, in *Truly Understood* (2008), Christopher Peacocke states that

[T]he extraordinarily rich and philosophically interesting epistemic phenomena exhibited by such an important concept as that of the first person can be explained by its fundamental reference rule. (77)

In Peacocke’s case, the proposed explicability of first person reference in terms of the token-reflexive rule is used as a case study for the more general thesis ‘that the fundamental reference rule for a concept contributes essentially to the explanation of the norms distinctive of that concept’ (77). Peacocke focuses on the self-consciousness of ‘I’ and its immunity to error through misidentification, and argues that these properties of first person reference can both be fully accounted for with reference to the token-reflexive rule.

The token-reflexive rule is primarily a means of accounting for the reflexivity of first person *reference*, since it is a rule dictating the behaviour of the first person pronoun. Nevertheless, insofar as first person thought also manifests reflexivity, and insofar as first person thoughts can be construed as the sorts of thoughts we express using the first person pronoun, the token-reflexive rule has also figured prominently in accounts of first person thought. Given the popularity of the token-reflexive rule within the first person literature, and the confidence surrounding its explanatory scope and potential, in the

¹⁰ For a more elaborate analysis of Campbell’s claim here, see Morgan (2014: 7-8).

present chapter I propose and assess a hypothetical account according to which the token-reflexive rule provides a sufficient analysis of first person thought. My purpose in fashioning such a category is largely heuristic: the aim is to determine whether a mere appeal to the token-reflexive rule, assuming it was made into a proper theory of first person thought, could possibly account for the entirety of the phenomenon. I think this is a possibility worth exploring since, if such a view is tenable, it will comprise a simple and highly economical account of the phenomenon.

This strategy also serves a diagnostic purpose and is largely motivated by a diagnostic ambition. In particular, it is designed to bring to the fore, and problematise, certain tendencies and assumptions which, in my opinion, dominate contemporary discussions of first person thought. One such tendency is to focus on the reflexivity of first person thought in accounting for it, whilst often overlooking its other properties. A quick look at the current literature shows that many discussions of first person thought revolve around trying to explain its guaranteed reference to the thinker; moreover, whether a given account succeeds in doing this is often construed as the main, if not the only, determinant of its success. This is exemplified, for instance, in philosophers' almost unanimous rejection of Perceptual Views on the grounds that they do not secure reflexivity. As we will see in the next chapter, despite this shortcoming, Perceptual Views may be well-equipped to explain other important dimensions of first person thought, like its cognitive significance. However, due to their presumed failure to vindicate the reflexivity of first person thought, such views are in most cases swiftly dismissed, and their possible merits ignored. This betrays a tendency to prioritise the reflexivity of first person thought over its other features in accounting for it, as well as a widespread tendency to reduce or assimilate first person thought to first person reference. For this reason, I refer to the cluster of accounts that manifest such tendencies as 'Reductionist Views' of first person thought. My aim in fashioning this category is to uncover these tendencies and assess their tenability when made explicit.

In what follows, then, I examine whether Reductionist Views could comprise adequate accounts of first person thought. I start at the beginning by asking whether, before its explanatory scope is expanded, the token-reflexive rule can fulfil its original explanatory purpose, namely that of explaining the reflexive and self-conscious reference of the 'I'.

2.2. What About Self-Consciousness?

2.2.1. The Insufficiency Worry and the Circularity Worry

As explained in the Introduction, it seems part of thinking first person thoughts and using the 'I' that the subject is aware, in doing so, that she is thinking about and referring to herself. Simple self-reference, i.e. reference to oneself without the knowledge that one self-refers, is not sufficient for first person reference or first person thought. For instance, and to take the linguistic first person as a case study, when Perry thinks 'The shopper with the torn sack is making a mess', he self-refers, since *he* is the shopper with the torn sack; however, his reference to himself is not first personal, because he does not know that it is himself he is referring to. By contrast, when Perry thinks '*I* am making a mess', he is aware that he is thinking about himself. It follows that first person reference is a distinctively self-conscious form of self-reference, and that, to capture the essence of this form of reference, an account must be able to explain both why the 'I' always refers to the one who uses it, and why it does so self-consciously. The same goes for first person thought.

A common, and potent, criticism against accounts of first person reference that restrict themselves to a simple appeal to the token-reflexive rule is that they fail to account for self-consciousness. The problem is that, in stating that any token of 'I' refers to the one who produced it, the token-reflexive rule may account for how an 'I'-user comes to self-refer, but it does nothing to explain the user's *awareness* that she self-refers. If this is correct, it follows that an appeal to the token-reflexive rule cannot on its own account for first person

reference, which is reflexive but also distinctively self-conscious. Let us call this the 'Insufficiency Worry'.

In *Self-Knowing Agents* (2007), Lucy O'Brien sums up the worry as follows:

A simple appeal to [the token-reflexive rule] is not sufficient to explain first-person reference because it has not been explained how a term of reflexive reference is expressive of self-consciousness. If first person reference involves not only a subject referring to herself, but also involves a subject self-consciously referring to herself then ... we have to invoke something further than [the token-reflexive rule] to explain first-person reference.' (2007: 59)

O'Brien is one of several theorists to have pointed out that a mere appeal to the token-reflexive rule cannot explain the self-consciousness of the 'I'. However, the problem with the token-reflexive rule is not just that it is silent with regards to the self-consciousness of the 'I'. It is also that any endeavour to modify it in order to capture the self-consciousness of the 'I' appears to spawn a vicious explanatory circularity. Let us call this second problem the 'Circularity Worry'.

Both the Insufficiency and the Circularity Worry are famously articulated by Elizabeth Anscombe (1981: 23):

[T]he 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 if that reflexive has in turn to be explained in terms of "I"; and if it is the ordinary reflexive, we are back at square one.

Anscombe assumes a slightly different formulation of the token-reflexive rule than the one we have been using here; on this formulation, the rule states that 'The word 'I' is a word that each speaker uses to refer to himself' (Bermudez 1998: 16). Anscombe's initial concern is that, so understood, the rule allows for subjects to use 'I' to refer to themselves without knowing that they are. This is because describing 'I' as 'The word each speaker uses to refer to himself' does not specify whether the form of self-reference delivered by the 'I' is self-conscious or not; and, as in Perry's shopper case, or Mach's shabby pedagogue

case, a subject can use a term to refer to himself without knowing that it is himself he is referring to. It follows from this that the token-reflexive rule fails to differentiate first person reference from simple self-reference, and thus to corroborate the self-consciousness of the 'I'.

The crux of the problem, as Anscombe sees it, lies in the reflexive 'himself' that the token-reflexive rule employs. Anscombe points out that the rule equivocates between two forms of the reflexive, each of which yields a different characterisation of the 'I'. One way in which the 'himself' can be construed is, to use Anscombe's terminology, as an 'ordinary', or 'direct', reflexive; that is, as a term that refers to the user, without the user knowing that it does (Anscombe 1981: 22). This is the form of the reflexive Mach employs, for instance, when he thinks 'He is a shabby pedagogue', without knowing that he is thinking about himself. Now, if the token-reflexive rule employs this form of the reflexive, it characterises 'I' as the word a subject uses to refer to himself, without knowing that he self-refers. However, in such a case, the rule describes first person reference as merely reflexive reference, and not as a self-conscious form of self-reference. It therefore fails to capture the self-consciousness of the former, and to differentiate it from simple self-reference. Hence the Insufficiency Worry.

As Anscombe sees it, to overcome the Insufficiency Worry, the token-reflexive rule must treat the relevant 'himself' as what she calls a 'peculiar', or 'indirect', reflexive; that is, as a reflexive that incorporates the subject's awareness that he self-refers, a *self-conscious* 'himself'. This is the form of the reflexive which we use to report other people's knowledge, beliefs, or claims about themselves (Castañeda 2001: 51); for instance, when we say 'Perry knows he has made a mess', or 'Florence Foster Jenkins believes she is a good singer.' What these propositions assert is that Perry knows that he, himself, has made a mess, and that Florence Jenkins believes, of herself, that she is a good singer. On this interpretation of the reflexive, when 'I' is characterised as 'The word each speaker uses to refer to himself', what is meant is that the 'I' is the word each speaker uses to refer to himself *self-consciously*, or *knowing it is himself he is*

referring to. Thus, if the token-reflexive rule assumes this form of the reflexive, it defines first person reference as a self-conscious form of self-reference, and successfully distinguishes it from simple self-reference. It thus overcomes the Insufficiency Worry.

Nevertheless, as Anscombe points out, this attempt to overcome the Insufficiency Worry puts the token-reflexive rule in even deeper trouble. As we have said, if the rule adopts the indirect form of the reflexive, it asserts that 'I' is the word one uses to refer to himself self-consciously, or knowing that it is himself he is referring to; however, the subject's self-consciousness, or knowledge that he self-refers, would seem itself to be first-personal, and thus to require mastery of the first person pronoun. Consider, again, the following propositions: 'Mach believes that he [that man] is a shabby pedagogue', and 'Mach believes that he [himself] is a shabby pedagogue'. The second, self-conscious use of 'he', in which Mach knows he is thinking about himself, is essentially the third-personal equivalent of the first person pronoun (Castañeda 2001: 73). It is what we, i.e. anyone other than Mach, use to report Mach's first-personal belief about himself, i.e. his belief 'I am a shabby pedagogue'. It follows that, as Anscombe puts it, 'that reflexive has in turn to be explained in terms of 'I'', and that grasping this form of the reflexive, and understanding its difference from the non-self-conscious 'himself', requires mastery of the first person pronoun. If this is correct, the indirect reflexive cannot be part of a rule which purports to explain the 'I', because then the rule will beg what it was intended to explain. Hence the Circularity Worry.

A champion of the token-reflexive rule could attempt to bypass the Circularity Worry by proposing a different formulation of the rule, on which the circularity may not arise (Bermudez 1998: 14-18). For instance, Campbell's articulation of the token-reflexive rule, as the rule that 'Any token of 'I' refers to the one who produced it', may seem to avoid circularity, because it does not employ the problematic reflexive (Bermudez 1998: 15). However, it is highly doubtful that this response could work. Let us think about a subject who uses 'I', and thus refers in accordance with the token-reflexive rule, in its Campbellian

formulation. In order to refer to himself, this subject would have to know that *he, himself* is the one who produced the relevant token of 'I', and this knowledge would be first personal (Bermudez 1998: 15-16). In other words, in order to secure reference to the user, the rule that 'Any token of 'I' refers to the one who produced it' would have to be coupled with the user's knowledge that *he* is the one who produced the relevant token. However, the knowledge that I, myself, have produced a certain token is first personal, and thus seems to require mastery of the first person pronoun. So we are back to the Circularity Worry.¹⁴

Thus the champion of the token-reflexive rule is confronted with a dilemma. On the one hand, if she takes the token-reflexive rule to employ the direct reflexive, she cannot make sense of the self-consciousness of first person reference, and cannot distinguish it from simple self-reference. On the other hand, if she tries to vindicate the self-consciousness of the 'I', by taking the rule to employ the indirect reflexive, she is faced with a circularity. Hence Anscombe's claim that '[T]he explanation of the word 'I' as the word which each of us uses to speak of himself is hardly an explanation.' If the token-reflexive rule employs the direct reflexive, then it is hardly an explanation of first person reference as the distinctively self-conscious phenomenon that it is. However, if it employs the indirect reflexive, it is hardly an explanation of the 'I' because it assumes what it tries to explain. What is more, attempting to reformulate the token-reflexive rule so that it both overcomes the Insufficiency Worry and escapes the Circularity Worry, for example by foregoing an employment of the problematic reflexive, seems unlikely to work.¹⁵

At the beginning of this chapter, I discussed the dominance of the token-reflexive rule within the first person literature and the confidence surrounding its explanatory potential, even with regards to first person thought. Following our examination, it appears that theories of first person reference and thought

¹⁴ Bermudez (1998: 14-18) attempts five different formulations of the token-reflexive rule, only to conclude that the circularity cannot be avoided.

¹⁵ A Reductionist View theorist could respond with a 'reductionist' strategy, which denies that first person reference is ineliminably self-conscious (e.g. Mellor 1989). This response is explored, and ultimately rejected, in O'Brien (2007: 59-65).

that aim to anchor their explanations of these phenomena entirely in the token-reflexive rule are very unlikely to work.¹⁶ The ensuing question is why the token-reflexive rule, given its limited explanatory potential, occupies such a dominant place in the first person literature, and why it is often invoked as a satisfactory analysis, not just of first person reference, but also of first person thought. Before addressing this question, I will make a brief but important detour to consider how the self-consciousness of first person reference (and first person thought) can be accounted for, given the Insufficiency Worry and the Circularity Worry.

2.2.2. Nonconceptual Self-Awareness

How *does* one account for the self-consciousness of first person reference? Importantly, that the challenge in doing so is faced not only by proponents of the token-reflexive rule, but by anyone who wishes to offer a philosophically tenable account of the 'I'. The challenge is summed up by Lucy O'Brien as follows:

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. (2007: 8)

¹⁶ Nonetheless, this should not be taken to suggest that first person theorists must entirely discontinue their appeals to the token-reflexive rule. It only suggests that our philosophical ambitions regarding the explanatory potential of the token-reflexive rule must be moderated. Despite its inadequacy in accounting for the self-consciousness of the 'I', the rule remains a highly reliable way of accounting for its reflexivity, and to this end can be appropriated by views of the first person that recognise both its powers and its limitations. Acknowledgment of this insight gives rise to 'hybrid' accounts of first person reference (O'Brien 2007), which incorporate the token-reflexive rule as a means of accounting for the reflexivity of 'I' whilst looking for other means for accounting for other features like its self-consciousness.

As we have seen, in theorising first person reference, one must account both for the fact that the 'I' refers to the one who uses it, and for the fact that 'I'-users *know* that they are self-referring. The latter demand raises an important challenge, however, because the knowledge that one self-refers appears to be first-personal, and thus to involve employment of the 'I'. This suggests that any account which sets out to explicate the 'I' will, at the point where it has to explain 'I'-users' knowledge that they self-refer, find itself having to beg it. Hence the challenge is to find a way to account for 'I'-users' knowledge that they self-refer without re-introducing the 'I' in one's account of the first person; i.e. to provide a non-circular explanation of how 'I'-users self-refer self-consciously.

Some philosophers have tried to block the problematic circularity by appealing to a primitive form of self-awareness which is meant to explain the 'I'-user's knowledge that she self-refers without invoking the 'I'. This form of self-awareness is conceived as *nonlinguistic* and *nonconceptual*, and thus as foregoing any employment or mastery of the first person pronoun; nevertheless, in being a kind of *self-awareness*, this mental state could be said to be first-personal, or at least *de se*. This alleged form of the first person is said to be more rudimentary than the conceptual or linguistic one, and to stand in certain important relations to the latter. For instance, it is said to cognitively underlie the conceptual or linguistic first person, such that every time a subject uses 'I' she also exploits, or expresses, the primitive form of self-awareness in which it consists. In addition, and very importantly, the conceptual or linguistic first person is thought to be, to some extent, *explanatorily dependent* on this form of awareness, such that appeal to the latter must comprise part of any complete account of the 'I'.¹⁸

How is appeal to a nonconceptual form of self-awareness meant to explain the self-consciousness of 'I'? This depends on how each theorist

¹⁸ This is where our recurrent distinction between first person *thought* and first person *reference* becomes significant: nonconceptual self-awareness can be seen as belonging to the realm of first person thought, which in turn is confirmed to be more primitive and fundamental than first person reference. Hence our assertion, at the very beginning of this thesis, that 'first person thought' may refer to diverse forms of first-personal cognition which may be linguistic or nonlinguistic, conceptual or nonconceptual.

conceives the relevant form of awareness, but the general form of the solution is the following. We have said that the circularity problem arises from the fact that any explanation of the 'I'-user's knowledge that she self-refers would seem to have to involve the 'I', since the knowledge that one self-refers is first-personal. The proponent of the nonconceptual *de se* argues that, although the knowledge that one self-refers may indeed be first-personal, an explanation of this knowledge need not involve the 'I'. This is because the subject's knowledge that she self-refers can be attributed to a primitive awareness she has of herself, which is nonlinguistic and nonconceptual, and which therefore does not involve, nor require, employment of the first person pronoun. In other words, the proposal is that, when a subject refers to herself using 'I', her knowledge that she self-refers is secured by a primitive form of self-awareness, which does not itself employ the 'I'; and since this form of self-awareness does not employ the 'I', appeal to it in explaining the subject's knowledge that she self-refers does not generate a circularity.

The precise nature and extent of the dependency of the 'I' on a nonconceptual *de se* is a matter of debate. Most important, for now, is the idea, endorsed by many proponents of the view at hand, that appeal to a primitive form of the first person is necessary if we are to explain the self-consciousness of first person reference. One place where this idea is put forth is Lucy O' Brien's *Self-Knowing Agents* (2007). Here, O'Brien invokes the nonconceptual *de se* as a solution to the Circularity Worry, particularly as the latter pertains to attempts to anchor an explanation of first person reference solely in the token-reflexive rule. After discussing Perceptual Views first, and concluding that these flounder when it comes to vindicating the reflexivity of the 'I', O' Brien resolves that the token-reflexive rule provides the best means of explaining reflexivity, and thus that it should, to this end, be made part of any account of first person reference. O'Brien then asks whether the token-reflexive rule could, by itself, account for the entirety of first person reference. After some reflection, which includes a discussion of Anscombe and the Insufficiency and Circularity Worries, O'Brien concludes that the token-reflexive rule could not possibly be sufficient to account

for the entirety of the 'I', because it cannot explain its self-consciousness without circularity. She therefore suggests that first person theorists must anchor their explanation of the self-consciousness of first person reference in some other feature of the phenomenon (58).

O'Brien's own proposal is that '[t]here is a non-conceptual awareness that is involved in first-person reference' (71), which explains the subject's knowledge that she self-refers when using 'I'. O'Brien calls this nonconceptual awareness 'agent's awareness', and theorises it as an awareness of our intentional actions. Her argument is that 'in the awareness an agent has of her own actions (agent's awareness) we find a primitive, non-conceptual, non-perceptual form of self-consciousness that [plays] a key role in explaining first-person reference' (73). Therefore, according to O'Brien, 'A subject who uses 'I' as governed by [the token-reflexive rule] will inevitably succeed in self-referring self-consciously, because of an awareness she has of her own thoughts and utterances' (57). The resulting account of first person reference is a 'hybrid' one, on which the 'I' is analysed partly in terms of the token-reflexive rule, and partly in terms of a primitive form of self-awareness (11).

Similarly, in *The Mirror of the World* (2014), Christopher Peacocke suggests that the basic form of self-awareness in which the nonconceptual first person consists is exhibited in conscious events such as seeing something as coming towards me, remembering an encounter, or having an action-awareness of moving my head (6). Such events involve self-representation, because their contents can be articulated using the first person pronoun, as in 'Something is coming towards *me*', '*I* had such-and-such encounter', or '*I* am moving *my* head' (6). However, Peacocke thinks that the contents of these mental states 'involve something less than the full conceptual first person content', since, for example, '[s]eeing something as coming towards one is something that can occur in subjects who lack concepts' (7). According to Peacocke, any account of the first person 'should recognize that there is first person content at the nonconceptual level' (1), which he also calls 'a nonconceptual *de se*.' Indeed, Peacocke distinguishes between three kinds of subjects, on the basis of their capacities for

self-representation: subjects that do not self-represent at all, subjects that employ the nonconceptual *de se*, and subjects that employ the first person concept (6). Crucially, Peacocke argues that ‘the nature of the first person concept is to be explained in part in terms of its relations to the nonconceptual *de se*’ (2014: 86), and that ‘Many of the distinctive phenomena involving conceptual forms of the first person can ... be understood theoretically only by relating them ultimately to this more primitive nonconceptual level’ (2).

I believe that the idea of a nonconceptual and nonlinguistic first person is explanatorily potent and philosophically appealing.¹⁹ In the remainder of this thesis, my efforts will be directed, both directly and indirectly, towards reinforcing its plausibility, and in Chapter 4, I will argue that an appeal to a nonconceptual first person may be able to supply an explanation of the Self-Concern Intuition. Nevertheless, skepticism about a nonconceptual first person may derive from certain commonly accepted assumptions about the nature of thought and its relation to language. These are the same assumptions which, as I will now go on to argue, motivate the category of Reductionist Views.

2.3. A Diagnosis

To sum up: the nonconceptual *de se* is understood as a rudimentary form of first person cognition, available even to beings who lack the capacity for conceptual or linguistic representation; a form of cognition that foregoes employment of the ‘I’, but which nevertheless constitutes a kind of *self*-consciousness, or *self*-representation. In operating below the level of language and concepts, this form of the first person belongs exclusively to the realm of thought, loosely construed, and can thus be conceptualised as a basic form of *first person thought*. In being wholly nonlinguistic and nonconceptual, this form of the first person does not warrant a semantic analysis. Moreover, according to

¹⁹ Another, very important champion of the nonconceptual first person is Bermudez (1998).

its champions, this nonconceptual form of self-consciousness is explanatorily prior to the conceptual and linguistic first person, i.e. the 'I'.

Skepticism about the notion of a nonconceptual form of first-person cognition, which is cognitively and philosophically prior to the conceptual first person, may derive from the fact that the understanding of the relation between thought and language that governs analytic philosophy postulates a reverse order of explanation:

Historically, the common principle uniting all the very diverse versions of analytical philosophy has been the priority of language over thought in the order of explanation. (Dummett 1996: 315)

Although early modern philosophy gave privilege to the mental, construing language as an instrument for communicating antecedently and independently formed thoughts, today thought is more commonly modelled after language, and thinking is conceived as a kind of 'inner saying' (Brandom 2000: 5). Dummett, one of the main proponents of the latter conception, articulates this contrast as follows:

We have opposed throughout the view of assertion as the expression of an interior act of judgment; judgment, rather, is the interiorization of the external act of assertion. (1973: 362)

The idea that 'judgment ... is the interiorization of the external act of assertion' is the idea that thought is language, represented at the level of mind. This entails that thought is fully semantically analysable, and that there is no part of thought, or no *kind* of thought, that lies beyond semantic analysis. It also entails that, to understand a given thought, it is not only necessary but also sufficient to understand its linguistic expression. Hence the explanatory priority of language over thought, asserted in the first passage by Dummett.

The claim that thought is modelled after language is a claim about the nature of thought, and its relation to language. The claim that language is explanatorily prior to thought is a claim about the philosophical methodology

one ought to follow in understanding thought. Dummett's own account of the relation between thought and language is anchored in his interpretation of Gottlob Frege, whom he construes as one of the main informants of what has sometimes been called the *linguistic turn* of analytic philosophy: a shift towards treating philosophical questions as quintessentially linguistic ones, and thus construing language as primary in the order of philosophical explanation. This shift is arguably synchronous with the development of analytic philosophy, and consequently lies at the very core of the latter.

I began this chapter by fashioning a category of Reductionist Views. These were understood to be views that focus on the referential questions surrounding first person thought, implicitly assimilating or reducing first person thought to its reference, and even presuming that first person thought can be adequately accounted for by appeal to the token-reflexive rule. If analytic philosophy is governed by the abovementioned understanding of the relation between thought and language, the philosophical motivations behind Reductionist Views should be clear. If first person theorists subscribe to the idea that thought is modelled after language, and abide by the ensuing methodological principle, it is only natural that they would treat first person thought as a primarily linguistic phenomenon. The associated focus on reflexivity and the token-reflexive rule is thereby also illumined. First, since the hallmark semantic feature of first person thought is its reflexivity, theorists who are motivated to approach the phenomenon from a linguistic perspective will naturally focus on this attribute. Secondly, given that the token-reflexive rule is widely believed to provide a sufficient explanation of reflexivity, theorists of this sort would naturally assume that the rule tells us all we need to know about first person thought. The Reductionist View could therefore be viewed as an application of the analytic understanding of thought.

Consequently, the notion of a nonconceptual first person recommends significant revisions to our understanding of first person thought. It also proposes to revise our dominant approach to first person reference. First, in suggesting that the 'I' involves a nonconceptual component, champions of this

view imply that a semantic analysis of first person reference is not sufficient for understanding the phenomenon. Secondly, in maintaining that the nonconceptual first person is explanatorily prior to the 'I', and that a proper understanding of the latter can only be obtained through a consideration of the former (in order to provide an explanation of self-consciousness, for instance), they turn the entrenched explanatory priority thesis on its head. On the ensuing view, it is not thought that must be understood through language, but the other way around.

For these reasons, the notion of a nonconceptual first person is likely to be met with some skepticism. After all, it is no accident that, although this notion has been circulating in the first person literature for a while, and despite its apparent explanatory advantages, it has yet to gain considerable traction. However, it cannot be rejected simply because it does not abide by our ruling philosophical prejudices, especially as it provides effective solutions to problems like the Circularity Worry, and, as I will go on to argue in Chapter 4, possible explanations of the Self-Concern Intuition.

A central premise of this thesis, as advanced in Chapter 1, is that our accounts of first person thought must take more care to accommodate the practical, affective, and normative dimensions of the phenomenon, and especially its association with self-concern. My discussion in this chapter has been aiming to illustrate that, in virtue of focusing largely on the referential dimensions of first person thought, the first person literature is far from according first person thought the treatment it requires. My aim has also been to critically understand the assumptions and tendencies motivating our current approach to first person thought, as a means of problematising this approach and recommending a more holistic one. This recommendation is made more pressing by the fact that, as I will now go on to suggest, a mere appeal to the token-reflexive rule is far from sufficient in vindicating the significance of first person thought and its connection with self-concern.²⁰

²⁰ A further diagnosis regarding our tendency to focus on the referential dimensions of first person thought will be provided in Chapter 4.

2.4. Significance and Self-Concern

As we explained in Chapter 1, first person thought exhibits a distinctive practical significance, which is evident in its influence on the subject's behaviour and action. It also exhibits a distinctively affective significance, consisting in its connection with a uniquely first personal class of emotions, and its link with self-concern. These are aspects of first person thought that any account of the phenomenon ought to explain. According to the Reductionist View, however, all that is involved in first person thought is reference to oneself in accordance with the token-reflexive rule. How could such a construal of first person thought explain its practical and affective significance and its connection with self-concern? The answer is simply that it could not. To say that all a subject does in first person thought is to refer in accordance with the token-reflexive rule is to say nothing about how, or why, first-personal thinking manifests any sort of practicality, affectivity, and normativity. It is also to provide no resources for such an explanation. In their preoccupation with the reflexivity of first person thought, Reductionist Views fail to account for or acknowledge these dimensions of the phenomenon.

If explaining the Self-Concern Intuition and the practical and affective significance of first person thought is a necessary requirement on any account of first person thought, as I have argued here, Reductionist Views do not comprise adequate accounts of the phenomenon. (And let us not forget that Reductionist Views are also inviable on more general grounds, e.g. with regards to explaining the self-consciousness of the 'I'. Indeed, it seems that the only feature of the 'I' they can explain is its reflexivity.) To understand the practical and affective significance of first person thought, and its connection with self-concern, we need to look beyond the token-reflexive rule. In particular, we need to search for an account that offers a more substantive characterisation of what we are doing when we represent ourselves in first person thought, and which provides the resources for explaining why doing so makes us act, feel, and care in special ways.

3. The Perceptual View

3.1. Introducing the View

The Perceptual View is so called because, at its core, lies the idea that first person thought functions on the basis of a broadly perceptual relation to the self. More specifically, proponents of the Perceptual View hold that subjects are in some way presented with, or aware of, themselves, and that a relation of self-presentation or self-awareness both underlies and facilitates their self-conscious thoughts about themselves. The Perceptual View also serves as a theory of first person reference, since the relation of self-awareness it appeals to can also be thought to facilitate, and be expressed in, our uses of 'I'.

In maintaining that first person thought and reference is based in perception, the Perceptual View models these phenomena after that of demonstrative reference. Demonstrative reference obtains in situations where the referring subject stands in an immediate perceptual relation to the object of reference, i.e. it is reference to objects in one's immediate perceptual environment. It is thought to involve usage of demonstrative pronouns like 'this' or 'that', and to be exemplified in propositions like 'This table is flimsy', or 'That man is a shabby pedagogue'. The standard way of accounting for demonstrative reference proceeds by fleshing out the kind of perceptual relation that holds between subject and object, and moreover, by accounting for the manner in which that perceptual relation supports the subject's thinking about or referring to the object. Suppose, for instance, that I am looking at a table in the room.

According to the Perceptual View, any thought about or reference to that table is facilitated by my visual relation to that object. That is, it is *by virtue of looking at the table* that I am able to represent it in thought and refer to it in language. This perceptual relation is meant to provide the basis for my reference to this object, and to enable my thoughts to latch on to it.

The Perceptual View faces various explanatory challenges. The central one lies in fleshing out the nature of the self-relation that first person thought and reference allegedly exploits. What is it to say that subjects are *related with*, or *presented with*, themselves? What does this presumed relation of self-relation, or self-presentation, consist in? In addition, since, on the Perceptual View, the subject is said to stand in a certain perceptual relation to her own self, proponents of the view must provide at least a rudimentary account of what this ‘self’ is that first person thought and the ‘I’ is meant to map on to. Different answers to these challenges have yielded different forms of the Perceptual View, which I will present in the following sections. In particular, I will discuss two prominent versions of the Perceptual View, each of which assumes a different account of the self. On the first one, the self is construed as the physical human being that each of us is; on the second one, the self is construed as an entity akin to a Cartesian Ego.

3.1.1. Evans’ Account

On one version of the Perceptual View, the ‘self’ is construed as a physical object, a body in the objective world, presented to the subject through her ordinary senses and other perceptual faculties like proprioception. This version of the Perceptual View is most prominently advocated by Gareth Evans in *The Varieties of Reference* (1982), where it is presented primarily as an account of first person reference.

Evans’ account of first person reference is modelled after his account of demonstrative reference, especially as this pertains to demonstrative pronouns. He therefore theorises ‘I’-thoughts in the same vein as ‘here’-thoughts and ‘this’-

thoughts (1982: 205). The distinctiveness of Evans' position consists in his construal of the 'self' as an element of the objective world, and specifically as the physical human being, or body, that each of us is. According to Evans, the latter comprises the object of first person reference, reference to which is facilitated by a complex perceptual relation that employs our five senses as well as our capacities for proprioception, kinaesthesia, and introspection.

How exactly do these perceptual faculties enable us to refer to ourselves, i.e. to the physical human beings that we are? Evans' answer to this question in part appeals to what he calls 'Russell's Principle':

Russell held the view that in order to be thinking about an object or to make a judgement about an object, one must *know which* object is in question—one must *know which* object it is that one is thinking about ... Russell took this Principle to require that someone who was in a position to think of an object must have a *discriminating conception* of that object—a conception which would enable the subject to distinguish that object from all other things. (1982: 65)

Russell's principle dictates that, in order to refer to an object, the subject must know *which* object it is; and that, to know which object it is, she must be able to identify or pick it out from all other objects. It follows that first person reference, construed as reference to the physical human being or body that one is, requires that the referring subject be able to identify herself as an object in the physical order of things, and, in particular, that she be able to distinguish her body from all other physical objects.

According to Evans, our perceptual faculties provide us with information and knowledge that enables us to identify ourselves as objects in the physical world, thereby making our reference to ourselves possible (220-224). For instance, our sense of touch makes us aware of our physical boundaries, therefore providing knowledge of where our own bodies end and others begin. Vision provides information about our general surroundings, and an awareness of our position relative to other objects; perceptual perspective gives us the necessary awareness of our position and orientation in space. According to

Evans, this information and knowledge that the subject receives about her physical self through her senses constitutes the means by which she is able to identify herself as the referent of 'I'. Hence the perceptual relation that fixes the reference of 'I' is essentially an information- and knowledge-providing relation. As he puts it, 'We clearly do have ways of gaining knowledge of ourselves, and I'-thoughts are thoughts which are controlled ... by information gained in these ways' (207). Although Evans' focus is on first person thought, understood mostly as thought in accordance with the first person pronoun, the same mechanism can be said to underlie and facilitate first personal thinking broadly construed.

Evans' account faces at least two substantial challenges, which have commonly been considered detrimental to its credibility. The first of these is famously put forth by Elisabeth Anscombe in 'The First Person' (1981: 31):

And now imagine that I get into a state of 'sensory deprivation'. Sight is cut off, and I am locally anaesthetized everywhere, perhaps floated in a tank of tepid water; I am unable to speak, or to touch any part of my body with any other. Now I tell myself "I won't let this happen again!" If the object meant by "I" is this body, this human being, then in these circumstances it won't be present to my senses; and how else can it be 'present to' me? But have I lost what I mean by "I"? 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.

On the account proposed by Evans, it is the information and knowledge provided by our senses that allows us to refer to ourselves first-personally. In the above passage, Anscombe articulates a situation in which a subject's perceptual faculties are entirely obstructed, and she is unable to obtain any knowledge or information about her physical self. If Evans is right, in such a situation, the subject should be unable to represent herself first-personally, since she has no access to the perceptual information that makes her thoughts about herself and her uses of 'I' possible.

As Anscombe points out – though not with reference to Evans himself – this is a highly unappealing result of Evans’ proposal. We are compelled to think that, even in the sensory deprivation tank, the subject has not ‘lost her self-consciousness’, nor ‘what she means by ‘I’’, and that she is still able to successfully employ the first person pronoun and comprehendingly represent herself in thought. In delivering the opposite result, Evans’ account ends up looking highly implausible. Moreover, in allowing for instances in which the ‘I’ fails to refer, Evans’ account fails to secure the guaranteed reflexivity of first person reference, which partly has to do with the ‘I’ being invulnerable to reference failure. This is also suggested by Anscombe at the end of the passage.

The second problem with Evans’ account is anticipated in the following passage by Armstrong (1984: 113):

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.

Imagine that, in a situation like this, a subject comes to entertain a thought that she would articulate through the proposition ‘I am touching my hair’. Clearly, the relevant thought would be false, since the subject would incorrectly identify a movement in another’s body as a movement in her own. Nevertheless, we would still be inclined to think that the subject’s use of ‘I’ in this situation successfully refers to herself. That is, the thought articulated in the proposition ‘I am touching my hair’ would not incorrectly refer to another subject – it would simply be false. However, in proposing that the reference of first person thought and the first person pronoun follows the source of the perceptual information provided to the referring subject, Evans’ account entails that, in a situation like this, such thought refers to the body from which the information is received. This points to a deep flaw in the account: in proposing that the reference of first person thought tracks the source of the perceptual information provided to the referring subject, it fails to guarantee its reflexivity in cases where the source of the information is distinct from the referring subject.

Hence Evans' account fails to vindicate the guaranteed reflexive reference of first person thought, by allowing, first, for cases in which it fails to refer, and secondly, for cases in which it refers to someone other than the user.²² The second version of the Perceptual View, which I will now proceed to consider, is designed to avoid such problems, by invariably guaranteeing the reference and reflexivity of first person thought. Nevertheless, as we will see, this may come at a different cost.

3.1.2. The Internal Perception Account

As mentioned at the beginning of this chapter, Perceptual Views construe first person thought as functioning on the basis of a perceptual relation to the self. This means that, in fleshing out the phenomenon, they must provide an account of the perceptual relation it supposedly exploits, as well as an account of the 'self' to which we are presumably perceptually related. As we have seen, Evans proposes that the 'self' is the physical human being or body each of us is, reference to which is accomplished through ordinary perception. By contrast, the second version of the Perceptual View construes the self as an object of thought, available to the subject via some form of internal perception. On the second version of the view, the object of first person thought is not an element of the objective order, i.e. a human body, but a self which, in a manner of speaking, lies within. On this view, subjects relate to and access this 'self' via some kind of internal perception, which also facilitates their first personal thinking and their uses of the first person pronoun.

This version of the Perceptual View long precedes Evans', and so does the conception of 'self' it assumes. The latter is rooted in the Cartesian *cogito*, on which the self is rendered as an object of thought necessarily available to the thinking subject through introspection. This so-called 'Cartesian Ego' has been definitively rejected on the basis of being metaphysically elusive. So has the

²² These problems with Evans' account are extensively and compellingly discussed in O'Brien (2007, Ch. 3) and Campbell (1995, Ch. 4), among others.

version of the Perceptual View at hand, both because it assumes this philosophically untenable conception of the self, and because it postulates a mysterious capacity of ‘inner sense’ through which subjects latch on to the Cartesian Ego.

Could not the view avoid such troubles by adopting a more plausible account of the self, on which the latter could be reached via our ordinary capacities for introspection? According to some, this form of the Perceptual View must assume a Cartesian Ego, if it is to ensure the guaranteed reflexivity of the ‘I’. This point is argued at length by Anscombe in ‘The First Person’. Anscombe begins by pointing out the requirement of guaranteed reflexivity, which any account of first person reference must fulfil. This requirement demands, first, that the ‘I’ always refers, and secondly, that it always refers to the one who uses it (30). Anscombe then proceeds to point out how difficult a requirement this is to meet, since, in order for the ‘I’ to always refer to the same object, this object must be constantly and without fail available (or as Anscombe puts it, ‘present’) to the subject, remaining identical across different uses of ‘I’:

Let us ... ask *only* how reference to the right object could be guaranteed ... It seems, then, that this reference could only be sure-fire if the referent of "I" were both freshly defined with each use of "I", and also remained in view so long as something was being taken to be *I*. Even so there is an assumption that something else does not surreptitiously take its place ... So we accept the assumption, and it seems to follow that what "I" stands for must be a Cartesian Ego. (30-31)

As hinted at the end of the passage, Anscombe believes that the guaranteed reflexivity requirement makes the postulation of a Cartesian Ego necessary, because the latter comprises the only object that could possibly play the relevant role. After the above passage, Anscombe asks us to ‘suppose that it is some other object’ that plays this role, like *this body* (31); then the deprivation tank case ensues. According to Anscombe, this case illustrates that there are instances in which our body is not present to us, yet we can successfully and comprehendingly use ‘I’ to refer to ourselves; therefore ‘this body’ could not

constitute the object of first person reference. She soon goes on to state, despairingly, that

Like considerations will operate for other suggestions. Nothing but a Cartesian Ego will serve. Or, rather, a stretch of one. (31)

The impasse she articulates is the following: either an account of first person reference vindicates its guaranteed reflexivity at the expense of postulating a Cartesian Ego, or it adopts a more plausible account of the self at the expense of failing the guaranteed reflexivity criterion. Faced with this impasse, Anscombe notoriously goes on to suggest that the 'I' does not refer at all.²³

3.2. Significance

Given this impasse, both forms of the Perceptual View have, for a while now, been brushed aside by first person theorists. Nevertheless, there is reason to think that such views may possess the tools to vindicate the cognitive significance of first person thought; namely, because they possess the tools to vindicate the cognitive significance of thought in general.

As we have seen, on the Perceptual View, first person thought and reference is modelled after demonstrative reference. As we have also seen, accounts of demonstrative reference proceed by fleshing out the perceptual relation that obtains between the subject and object of reference, and by explaining the manner in which this relation supports the subject's thought about and reference to the object. The same perceptual relation can also be appealed to to explain the practical significance of thought. That is, it is possible to ground the practical attitudes and dispositions we exhibit towards an object (including the way we care about it) in the perceptual relation on the basis of which we represent that object. This is a quintessentially Fregean insight,

²³ It is precisely in an attempt to avoid the need for a Cartesian Ego that Evans proposed an account of first person reference that construes the self a body in the physical order of things (Evans 1982: 256-257).

alluding to the idea that the cognitive significance of a given thought is determined, not only by its reference, but also by its *sense*. Put plainly, the idea is that there are different ways of being presented with a referent (i.e. different ‘senses’), and that the manner in which one is presented with a referent largely determines her dispositions and behaviour towards it.

For example, it is generally acknowledged that considerations about the subject’s *spatial* relation to the object (e.g. their relative distance) can play an important role in explaining the subject’s dispositions towards that object. This fact fits nicely within accounts of demonstrative reference. For example, Evans, who works within a Fregean framework of reference, has the following to say about ‘here’-thoughts, in which he is particularly interested:

It is difficult to see how we could credit a subject with a thought about *here* if he did not appreciate the relevance of any perceptions he might have to the truth-value and consequences of the thought, and did not realize its implications for action (consider, for instance, a thought like 'There's a fire here'). (1982: 161-162)

This passage makes two claims. First and foremost, that thoughts conditioned by a perceptual relation between subject and object have certain practical manifestations (or ‘consequences’) and implications for action. Take a subject who spots a fire in his vicinity, e.g. by seeing the flames, or smelling the smoke, or feeling abnormally hot, and who, on the basis of such perceptual information, articulates the thought ‘There’s a fire here’. Clearly, this thought would have to influence this subject’s actions and his deliberative and affective apparatus in some way. He would most likely start to feel fearful and uneasy, beginning to deliberate about his options for escape. Sooner or later, he would probably take action, e.g. by evacuating the area and notifying the fire department.

Importantly, it is the subject’s representing the fire as being *here*, on the basis of his specific perceptual input, that determines the exact nature of his reactions. This becomes clear when we think about how the subject’s responses (and thoughts) may differ in response to similar, but different, perceptual input. If the subject had come to think ‘There’s a fire in Texas’ on the basis of seeing it

on television, he would have little cause to feel either terror or fear. By contrast, if the subject felt a burning sensation and looked down to see that his trousers were on fire, he would experience panic and terror rather than fear or unease, and he would scream for help instead of calling the fire department. In each of these cases, the subject's behaviour can be understood as a function of the particular perceptual input he receives. Moreover, the subject's spatial relation to the object of reference could receive a more detailed characterisation, e.g. through an articulation of its directional properties, which would also yield a more nuanced explanation of his behaviour. A distant growl coming *from the west* would dispose a subject to move *towards the east*; a growl a few feet away *to his right* would dispose him to run *to the left*.

According to Evans, the disposition to exhibit such reactions is part and parcel of entertaining 'here'-thoughts, and generally thoughts based on perception. Indeed, his second claim in the above passage is that it is hard to imagine a subject thinking 'There's a fire here' without being disposed to exhibit the relevant reactions, or at least appreciating that he should be so disposed. But Evans also says the following:

[W]e must say that having spatially significant perceptual information consists at least partly in being disposed to do various things ... The subject hears the sound as coming from such-and-such a position, but how is the position to be specified? Presumably in *egocentric terms* (he hears the sound as up, or down, to the right or to the left, in front or behind). These terms specify the position of the sound in relation to the observer's own body; and they derive their meaning in part from their complicated connections with the subject's *actions*. (155)

Thus Evans makes the much bolder claim that to represent something as being in a certain position in space *consists in* being disposed to act in certain ways towards it, at least partly. For instance, according to Evans, to represent a sound as coming from my left is, in part, to be disposed to turn my head to the left in response to the sound (155). For Evans, then, our very understanding and representation of space is dispositional, and practicality is embedded in the conceptual structure of perceptual thinking.

Here it is not my concern to evaluate Evans' claims, but simply to illustrate some of the ways in which demonstrative reference accounts could vindicate the practical and affective significance of thought. While there is certainly much more to be said about how exactly the perceptual relation involved in demonstrative reference explains a subject's reactions to her thoughts, and the connection of such thoughts with action, the above remarks illustrate, at least, that demonstrative reference accounts supply a promising framework for doing so.

The central argument of this thesis, as presented in Chapter 1, is that explaining the Self-Concern Intuition and the cognitive significance of first person thought should neither be optional nor an afterthought. Instead, it should comprise an essential requirement on theories of first person thought. Now, if it is true that Perceptual Views may be able to fulfil this requirement, this is enough to justify – or, indeed, require – a more careful consideration of such views. In the remainder of this thesis, my task will be to determine whether we can obtain an explanation of the Self-Concern Intuition on an account of first person thought that proceeds along the lines of a Perceptual View. Of course, before this can be done, it must first be determined whether a version of the view can be articulated that gets off the ground. Thus I begin, in the next chapter, by proposing an account of first person thought, or a structure for one, which proceeds along the lines of a Perceptual View while attempting to avoid the problems that plague Evans' account and the internal perception view. The proposed account will be articulated through reflection on the problem of other minds, and it will be called the 'Immediate Access' account of first person thought, because it will construe first person thought as resting on the subject's immediate access to her mind and its contents.

Before I proceed, it is worth noting how the insights presented so far in this chapter bear on those of Chapter 2. There, I diagnosed a general tendency to focus on the referential questions surrounding first person thought, whilst overlooking its other features and perhaps even assimilating it to first person reference. I also proposed that this tendency is exemplified in theorists'

dismissal of Perceptual Views on the grounds that they do not adequately address the referential questions surrounding first person thought. Now the reasoning behind this proposal becomes apparent. If there is any reason to think that Perceptual Views may be able to explain the cognitive significance of first person thought and its connection with self-concern, this is a significant explanatory advantage, and enough reason to try to make them work, or at least exploit their explanatory resources. That Perceptual Views are rarely considered as potential candidates for an account of first person thought confirms that the demand to accommodate the non-referential aspects of the phenomenon is not sufficiently appreciated. Therefore, my forthcoming endeavour to determine whether an account along the lines of a Perceptual View may be possible, and whether, in that case, it may be able to accommodate the Self-Concern Intuition, marks a start at addressing and correcting these tendencies.

4. The Immediate Access View

4.1. Introducing the View

4.1.1. The Problem of Other Minds

According to Perceptual Views, subjects are somehow presented to or aware of themselves, and a relation of self-presentation or self-awareness facilitates their self-conscious thought about themselves. Evans's account and the internal perception account each proposed a different understanding of the self-awareness that underlies first person thought, corresponding to different construals of the 'self'; yet neither view turned out to be successful. A good place to start in proposing a different account of first person thought along the lines of a Perceptual View is by proposing a different understanding of the self-relation or self-awareness that first person thought exploits; and a good place to start in understanding the way in which we relate to, and are aware of, ourselves, is by comparing it with the way in which we relate to, and are aware of, others. Such a comparison can be found in *the problem of other minds*.

Analytic philosophy of mind has always assumed a certain asymmetry between our awareness of ourselves and our awareness of others. This asymmetry is thought to follow from the fact that we each possess an immediate awareness, or knowledge, of our own minds, which we lack with respect to the minds of others. Moreover, and importantly, the presumed fact that we can only be indirectly acquainted with other minds is thought to raise a problem regarding our capacity to know the latter accurately, or even at all. This problem

can take several forms, which, together, comprise what is traditionally known as *the problem of other minds*.

One form that the problem of other minds can take concerns whether, and how accurately, we can know the specific contents of other people's minds, i.e. what other people think, feel, desire, believe, etc. Where 'knowledge of minds' is taken to mean 'knowledge of the *contents* of minds', the asymmetry between our knowledge of ourselves and our knowledge of others is thought to spring from the fact that, on the face of it at least, we are in a much better position to know what we ourselves think, feel, desire, believe, and so on, than we are to know these facts about others. This is because, whereas we are able to know most of our own mental states directly, simply by virtue of experiencing them, to gain knowledge of other people's mental states we must necessarily engage in observation of their outward behaviour, and subsequent observation-based inference.

Consider, for example, situations in which you felt intensely angry. Presumably, to *know* that you were angry, you did not have to ask yourself whether you were angry, or check the mirror to determine whether you were wearing an angered expression. You knew that you were angry simply by virtue of *feeling* angry; we may say that you had an *immediate awareness* of your anger, a direct and non-inferential knowledge of your mental state. By contrast, since you cannot feel other people's feelings directly, as you do your own, to know that someone else is angry, you must rely on inferring their mental states from their testimony or behaviour. This means that you could only know if someone else was angry if they told you so, or if you heard their aggravated yells or observed their flushed faces. However, since people's overt behaviour is not always the best guide to their subjective states, we can never be certain about our approximations of the contents of other people's minds; and neither is our knowledge of others' mental states as frequent and rich as our knowledge of our own. Hence the first form of the problem of other minds.

Here we must mention that some commentators have questioned whether this form of the asymmetry obtains. This is usually done by

undermining the presumed directness or accuracy of self-knowledge, and sometimes even by trying to assimilate the latter to our knowledge of others. Most famously, Gilbert Ryle (1949) proposes that, just like our knowledge of others' thoughts, our knowledge of our own thoughts is achieved inferentially, and thus indirectly. The same proposal has been forcefully argued for, more recently, by Quassim Cassam (2017). Nevertheless, none of these efforts have been successful in undermining the intuition, centrally underlying the problem of other minds, that our access to our own mental states differs in some way from our access to other people's, and that it exhibits at least *some* form of 'directness' that the latter does not. While claims to the effect that our knowledge of ourselves is more accurate or somehow better than our knowledge of others may be minimised, and while it may be conceded that some forms of self-knowledge are indeed achieved indirectly, the intuition of asymmetry is difficult to do away with. As Richard Moran puts it (2001: xxix-xxxi),

This basic thought of immediacy does not claim, for instance, that the mind's access to itself is infallible, or complete, or can't be corrected by others or by external evidence... Further, it is not argued here that all of a person's awareness of his mental life is achievable in this "immediate" way; much of the hard-won knowledge of oneself will be based on the same kinds of considerations available to others and fraught with the same possibilities for error and misinterpretation ... But at the same time, it is equally natural, and I think unavoidable, to think that, for a range of central cases, whatever knowledge of oneself may be, it is a very different thing from the knowledge of others.

A second form of the problem of other minds does not ask whether we can know what goes on in other people's minds, but whether we can know that others have minds at all. Thus, whereas the previous construal of the problem simply assumed that others are minded, i.e. that they have thoughts, feelings, etc., and questioned how far we can know *what* they think and feel, the second construal asks whether, and how, we can know that others think and feel at all. The worry here is that, unless we can 'enter' other people's minds and experience these minds directly, as we do our own, we can never be certain that they have

inner lives in the first place. Thus, on this formulation of the problem, the asymmetry between our awareness of our own minds and our awareness of the minds of others threatens skepticism about the very existence of the latter.

Of course, we do assume that others are minded, and live our lives accordingly. Others act and behave exactly as if they have minds – so why should this not be enough for us to conclude that they do? What reason do we have to think that the signs of mindedness other people outwardly exhibit do *not* correspond to, and are *not* caused by, inner experiences? These are powerful remarks, but the sceptic's point is even more powerful: however natural it may feel to assume that others are minded, the existence of other minds is ultimately impossible to confirm. *Unless* I can be directly aware of the minds of others, like I am of my own, there is no way I can know that others are minded. This is a quintessentially Cartesian insight: part of the point of 'I think, therefore I am' is that it only applies to the first person.

There are more strands to the problem of other minds, but only these two need to be considered for my current purposes.²⁴ It is important to reiterate that both of these issues are rooted in the presumed asymmetry between our awareness of our own minds and our awareness of the minds of others, which is due to the former exhibiting a certain *directness* that the latter does not. First, we are in a better position to know the contents of our own minds because we are, to some extent at least, directly aware of them. Secondly, we are in a position to be certain that we are minded because we are directly aware of our own mindedness. These two claims of directness are distinct. While the first is a claim about our (direct) awareness of our *mental states*, the second is a more general claim, about our (direct) awareness of our *minds*. And while the latter is quite difficult to articulate, the intuition supporting it is it nevertheless easy to grasp, and highly compelling. The idea is that each subject possesses a distinctive,

²⁴ The third is a *conceptual* problem, which asks whether we can even make sense of the concept of a mind that is not our own. It rests on the idea that, since our concept of 'mind' is derived exclusively from our experience of our own minds, and signifies a centre of consciousness or viewpoint that is strictly first personal, the concept of a centre of consciousness that is not ours may be unintelligible.

unmediated awareness of her own mind, or inner world, or consciousness, or self; an awareness she does not, and probably could not, have with respect to others; and which provides her with a unique certainty of her own existence. This is the same intuition that underlies the Cartesian *cogito*, as well as the one expressed in Frege's claim that in first person thought, a subject is 'presented to himself in a particular and primitive way, in which he is presented to no-one else.'

4.1.2. The View

Now, let us take stock. We embarked on a discussion of the problem of other minds as a way into articulating the relation of self-awareness that grounds first person thought, according to Perceptual Views. We are now in a position to distinguish at least some properties of the relevant relation. First, a subject's relation to herself involves a position of epistemic privilege with regards to her own mental states. Secondly, the relevant self-relation incorporates a certainty of one's own mindedness. And thirdly, the subject's awareness of herself and her mental states is peculiarly direct. Putting all these together, we can characterise the relation that grounds first person thought as a relation of immediate awareness of, or access to, one's mind and its contents. The ensuing account construes first person thought as exploiting this exact relation, and it can therefore be called the 'Immediate Access View' of first person thought.

Ultimately, the insight I am pursuing in theorising first person thought through the problem of other minds is the following. We have, on the one hand, the literature on the first person, which observes and tries to make sense of our self-conscious thinking about ourselves. We also have some articulations of a Perceptual View of first person thought, which are promising in some respects, but unsuccessful in others. On the other hand, we have a vast literature on the problem of other minds, which contains important insights about a subject's relation to and awareness of herself. These insights are not wholly uncontroversial, but they are nevertheless widely acceded to; most philosophers

agree that subjects possess at least some sort of special awareness of their own minds and mental states. I am suggesting that we can benefit from bringing these two literatures together, and that the problem of other minds could hold the key to articulating a more plausible account of first person thought along the lines of a Perceptual View.

With this in mind, let us say a little bit more about how the suggested account of first person thought could proceed. I will begin by looking at Kieran Setiya's 'Selfish Reasons' (2015), in which he argues for an account of first person thought as based on our immediate knowledge of our mental states. Setiya's interest in the paper is in the connection between first person thought and self-concern, but from the viewpoint of justification rather than explanation. To make sense of this connection, Setiya sets out to provide an account of first person thought along the lines of a Perceptual View, by 'investigat[ing] the relation to myself that is involved in first person thought' (457).²⁶ He begins by pointing out that first person thought refers to the thinker (457-458), but says that this could not be sufficient to characterise it, because the token-reflexive criterion says 'nothing about the grounds on which you apply [the 'I'] or form beliefs involving it' (458). He subsequently asks us to imagine a subject who comes to know her own beliefs in the same way as she comes to know other people's beliefs, i.e. through inference or testimony. This is a subject who, in thinking that *p* is true, does not immediately know that she believes that *p*. Rather, to achieve such knowledge, she has to ask herself whether she believes that *p*, or infer whether she believes that *p* from her other beliefs or behaviour.

According to Setiya, such a subject could not be thought to possess a proper grasp of the first person concept. As Evans compellingly argued, the relation in which I stand to my own beliefs involves a capacity to know that I believe that *p* simply by taking *p* to be true (Evans 1982: 250); this is a capacity for immediate, non-inferential, non-testimonial knowledge of my beliefs based

²⁶ Setiya makes his Fregean sympathies clear throughout the paper.

on my representations of the world. According to Setiya, this capacity is a necessary component of, and indeed condition for, first person thought:

If you could have a concept that meets our criteria, yet be unable to tell that you believe that p when you believe that p , it is not a concept of the first person. If this is not in fact possible, that is because possession of such a concept requires the capacity for self-knowledge ... Either way, the capacity for self-knowledge is a condition of first person thought. The relation by which I refer to myself in the first person is not simply that of being the thinker of these thoughts, but being the object of immediate knowledge. The first person concept refers to the one whose thoughts can be known in this way. (459)

Therefore, according to Setiya, a subject who uses 'I' but lacks the capacity for immediate knowledge of her own beliefs is either impossible, or the concept she is using is not really that of the first person.

On the basis of these observations, Setiya concludes that first person thought refers to the object of 'psychological self-knowledge', i.e. the subject whose mental states can be known non-inferentially and non-testimonially. He then captures the self-relation involved in first person thinking through the following thesis (461):

IMMEDIATE KNOWLEDGE: When I think of myself in the first person, I do so in virtue of standing to myself as the object of immediate knowledge, knowledge that is non-inferential, non-testimonial, and immune to error through misidentification relative to the concept of myself that figures in the self-ascription of beliefs.

On the relevant account, first person thought is immune to error through misidentification because I cannot, in thinking 'I believe that p ', be mistaken that it is I who believes that p . This follows from the manner in which I am aware of my own mental states, and specifically from the fact that there is no one else whose mental states I can know in this particular way, i.e. non-inferentially. Hence the reflexivity of first person thought is also guaranteed.

Setiya also considers how an analysis of the first person thought as based on an immediate awareness of one's mental states interacts with Evans' account,

according to which the reference of 'I' is primarily fixed through proprioceptive awareness, kinaesthetic awareness, perceptual perspective, etc. As Setiya correctly points out, the proposed analysis is not incompatible with Evans' account, and, indeed, the two can be conjoined.²⁷ According to Setiya, psychological self-knowledge is distinctive in being both essential to and sufficient for first person thought (460, 461), but it 'does not play an exclusive role in fixing its reference' (461). Rather, '[t]he referent of first person thought is the object of immediate knowledge in all its forms' (461); that is, it can be construed both as the object of immediate knowledge of one's mental states, and as the object of knowledge gained through proprioception, kinaesthesia, etc. (459-462). I agree that acknowledging the problems with Evans' account does not mean that we should disregard its insights. Indeed, insofar as we have proposed a more secure means of fixing the reference of first person thought, we may still incorporate Evans' insights into our account of the phenomenon. As Lucy O'Brien also puts it, 'Denying that perceptual information plays a critical role in explaining how first-person thought reaches its referent does not imply that it may nevertheless be a broader necessary condition of such thought' (2007: 38).

A final question concerns how this part of the Immediate Access View interacts with the internal perception account. One may worry that, just like the latter, it also postulates a Cartesian Ego, as the subject or owner of the mental states of which we have immediate knowledge. However, to claim that subjects are introspectively aware of their mental states is not to claim that they are introspectively aware of a subject 'having' those states. The Immediate Access View assumes no more than the standard philosophical account of self-knowledge, which is non-committal with regards to the awareness or existence of a subject of experience. This applies to Setiya's view as well, and Setiya fends off the relevant objection through his rejection of Mark Johnston's claim that introspection yields awareness of 'an arena of presence

²⁷ Evans himself also incorporates psychological self-knowledge in his account of first person reference (1982: 224-235), despite this not comprising his principal focus.

and action’, a ‘virtual frame’ or ‘container’ of our mental states (Johnston 2010: 139-140), at the centre of which lies a self (Johnston 2010: 192-199). Echoing Hume, Setiya proposes that ‘when I introspect, all I find are facts about me in relation to the world, not a virtual frame or container for mental stuff’ (465).

Now let us consider the second half of the Immediate Access View, according to which first person thought exploits an immediate awareness of one’s own mind, as distinguished from one’s particular mental states. What is this ‘awareness of mind’, and how is its object to be construed? It may be thought that, to make sense of such a proposal, we must resort to a Cartesian Ego – how else could we comprehend this elusive ‘mind’, or ‘self’, of which we are allegedly directly aware? Instead, I would like to suggest that we can appeal to the notion of a nonconceptual and nonlinguistic form of self-awareness, already introduced in Chapter 2. In designating an elemental form of self-awareness, the nonconceptual and nonlinguistic first person suggests a convenient and plausible construal of the immediate awareness of ‘self’, or ‘mind’, on which first person thought is partly based, according to our account. This way of understanding the relevant form of self-awareness avoids the postulation of a Cartesian Ego. It also does not preclude the possibility that first person thought incorporates bodily self-awareness, as per Evans.

4.1.3. Some Notes

According to the Immediate Access View, then, first person thought is based partly on our immediate knowledge of our mental states, and partly on a nonconceptual and nonlinguistic form of self-awareness. It is important to note that, thus articulated, the Immediate Access View is best seen as providing a structure on which a fully-fledged account of first person thought can be built, rather than as a fully-fledged account in itself. Questions remain to be answered, and particulars to be defined.

One parameter along which the Immediate Access View can be defined further concerns how the nonconceptual self-awareness involved in first person

thought is to be understood. Currently, the literature contains three main ways of understanding the relevant awareness: first, as a pre-reflective form of awareness that qualifies as an awareness of self without, however, positing the latter as its object; secondly, as a sense of ‘ownership’ or ‘mineness’ that accompanies our conscious experiences; thirdly, as an awareness or perception of the self as the subject of experience, variously construed such that it avoids commitment to a Cartesian Ego (Smith 2020). Each of these interpretations would presumably yield a different version of the Immediate Access View, corresponding to different possibilities for explaining the Self-Concern Intuition. For instance, as already mentioned earlier in this thesis, and as we will see in more detail later on in this chapter, Lucy O’Brien suggests an understanding of nonconceptual self-awareness as an of *agency*, which is essentially concerned due to self-concern comprising a constitutive component of agency. The interpretation I am most sympathetic to is also the first one; that is, when I speak of nonconceptual and nonlinguistic self-awareness, I mostly have in mind a pre-reflective form of self-consciousness underlying our conscious experiences without, however, taking the self as its object.

A second parameter concerns how the two components of the Immediate Access View – immediate self-knowledge and nonconceptual self-awareness – may be thought to interact, and the role each plays in first person thought. One question, for example, is how each figures in the subject’s first-personal cognition and psychology. For instance, in keeping with my understanding of nonconceptual self-awareness as a pre-reflective form of self-awareness, I imagine such awareness as enduringly present in all of conscious experience; by contrast, I take self-knowledge, understood as knowledge of one’s particular mental states, to be episodic, occurring through distinct and disparate acts of introspection.

A third parameter concerns the role that perception plays in the Immediate Access View, and, consequently, whether the account qualifies as a Perceptual View of first person thought. First, is our capacity for immediate self-knowledge, most commonly rendered in terms of introspection, a perceptual

capacity? Secondly, is nonconceptual self-awareness a perceptual form of awareness, and does it involve anything, such as a self, being perceived? These are complex questions, and how one answers them will largely depend on the stance one takes towards the other two parameters. For instance, there are certain interpretations of nonconceptual self-awareness on which it does qualify as perceptual, particularly those that amount or harken back to internal perception accounts. On my understanding of nonconceptual self-awareness as a pre-reflective form of self-consciousness, the relevant capacity is not perceptual. On the other hand, whether introspection is a perceptual (or quasi-perceptual) capacity is a difficult and highly contested question, which I cannot purport to answer here;²⁸ so the possibility remains that at least one half of my account might appeal to perception. The upshot is that an Immediate Access View may or may not qualify as a Perceptual View, depending on how a series of important questions are answered. At any rate, as illustrated by my argumentative strategy in this chapter, Immediate Access Views do proceed along the lines of Perceptual Views, insofar as they purport to theorise first person thought primarily through a relation of self-awareness.

While the Immediate Access View must be developed and examined in much more detail than I have done here, on this preliminary formulation, it suggests reasonable and *prima facie* reliable ways of accommodating the central properties of first person thought. First, in relying on ways of fixing the reference of first person thought which are only available to the thinking subject, it guarantees its reflexivity and immunity to error through misidentification. Secondly, as discussed in Chapter 2, nonconceptual self-awareness, which the proposed account incorporates, suggests a non-circular way of accounting for the self-consciousness of first person thought. The question now remaining to be answered is whether such views can accommodate the cognitive significance of first person thought, and particularly its connection with self-concern.

²⁸ For instance, Morgan (2014) suggests an account of first person thought as (partly) based on introspection, which he presents as a perceptual account.

4.2. Explaining the Self-Concern Intuition

4.2.1. Immediate Self-Knowledge

According to the Immediate Access View, first person thought exploits our immediate, non-inferential knowledge of our mental states, and a direct, nonconceptual, pre-reflective form of self-awareness. On the strategy pursued by Perceptual Views, after which Immediate Access Views are modelled, the cognitive significance of thought is located in the relation through which subjects come to represent their objects in thought. This suggests the possibility of explaining subjects' distinctive concern for themselves, as manifested in first person thinking, in terms of the relation of immediate self-knowledge or immediate self-awareness that supports such thinking, according to the Immediate Access View.

Let us begin by considering whether the Self-Concern Intuition may be explicable in terms of the subject's immediate knowledge of her mental states. Could it be that subjects care about themselves in a special way because they know – at least some of – their mental states directly, i.e. without inference? Could self-concern somehow be a consequence of epistemic immediacy? In 'Selfish Reasons', Kieran Setiya dismisses the possibility almost immediately (2015: 467):

Why care so much about the one you know first-hand, without the need for inference, whose beliefs you can access in a special way? The epistemic relation that secures first person thought is not a basis for special concern any more than the relation of speaking this utterance or thinking this thought.

Though Setiya is interested in the justification rather than the explanation of the Self-Concern Intuition, his complaint also extends to the latter case. Why would one's knowing something non-inferentially make one care about the thing known in a special way, whatever that may be? Some mathematical propositions are also thought to be known without inference, yet we do not care about them

in a special way – indeed, we do not care about them at all.³¹ As Setiya points out, it seems that an understanding of first person thought as resting on immediate self-knowledge is in no better position to vindicate the Self-Concern Intuition than an account on which it is simply thought in accordance with the token-reflexive rule, as per the Reductionist View of Chapter 2. Thus Setiya, who construes first person thought primarily in terms of epistemic immediacy, arrives at the conclusion that self-concern cannot be justified first-personally.

Setiya's goal in 'Selfish Reasons' is to formulate an account of first person thought on which self-concern can be justified, just as my goal here is to propose an account on which it can be explained. Both here and there, the requirement to accommodate self-concern is prioritised in providing an account of first person thought. Therefore, as O'Brien and Guillot point out in 'Self Matters' (unpublished), it is strange that Setiya would choose to propose an account of first person thought in terms of immediate self-knowledge, only to then proclaim that it fails his own justification requirement. It may very well be that, at the end of the day, self-concern cannot be vindicated on any account of first person thought, but more effort is due before such a conclusion can be reached. Here, I will do my best to fulfil my primary explanatory objective, and although my Immediate Access View theorises first person thought not just in terms of immediate self-knowledge, but also in terms of nonconceptual self-awareness, I will spend some more time trying to determine whether self-concern could possibly be explained in terms of epistemic immediacy.

In the first person case, epistemic immediacy entails epistemic abundance. Our capacity to know our mental states directly, and the consistent availability of our mental world to introspection, means that we can, and do, know many things about ourselves. There is also no doubt that our knowledge of ourselves is considerably, if not immeasurably, greater in degree than our knowledge of others. Could an explanation of the Self-Concern Intuition appeal to epistemic abundance? Could it be that we care about ourselves in a special way because we know so much about ourselves?

³¹ I am grateful to Professor Mark Textor and Dr. Lea Salje for pointing this out.

It is not extravagant to think that the more we know about someone the more likely we are to care about them. The potential link between knowledge and concern is often addressed, for instance, in philosophical and psychological discussions of empathy. Knowledge of another's mental states and circumstances is commonly construed as a necessary condition for empathy,³² and empathy is commonly thought to induce, or to at least enjoy a strong association with, concern, compassion, sympathy, helping behaviour, and even altruism.³³ What is more, in common sense, knowing another's mental states, empathy, and concern are often equated with one another (Batson 2009: 3-8). Hence the widespread conviction that knowledge of others can bar judgment, heal division, and foster sympathy and concern, both on a personal and on a social level.³⁴

Now, if knowledge of another's mental states leads or amounts to empathy, and empathy leads or amounts to concern, this suggests a plausible link between knowing a subject's mental states and caring about that subject. This link could be appealed to to explain why epistemic abundance might lead to special self-concern in at least two ways. First, one could explain self-concern *through* empathy, arguing that subjects' abundant knowledge about themselves facilitates an attitude of empathy towards self that is distinctive in its scope and success, and which in turn fosters a distinctive attitude of self-concern. Secondly, and less ambitiously, one could take the above remarks as evidence of a general association between knowing and caring, meant to substantiate the proposal that epistemic abundance facilitates concern.

As far-fetched as these explanatory possibilities may seem, I bring them up because I do believe that our extensive knowledge of ourselves might have something to contribute to our self-concern. That said, an appeal to epistemic

³² See, for example, Sober and Wilson (1998: 234-235), Snow (2000), Rogers (1975: 3), and Batson (2009: 4).

³³ See, for example, Batson (2010), Hoffman (2008), and Toi & Batson (1982).

³⁴ This conviction also find an application beyond the sphere of interpersonal relationships. Knowledge is thought to increase concern about all sorts of issues – political, environmental, etc. The very notion of an 'awareness campaign' rests on the presumed association between knowing more and caring more.

abundance is clearly too weak to serve as the primary explanation of the Self-Concern Intuition. First, the purported association between knowledge and concern, on which the argument rests, is contingent. Even if knowledge is associated with concern, this link is not sure-fire, and knowing many facts about someone while failing to be concerned about them is always possible. If we want to understand self-concern as a built-in feature of first person thought, we need something considerably stronger than mere association backed up by experience, intuition, or experiment to serve as its explanation. Secondly, the purported explanation models self-concern after concern for others, thus failing to vindicate its distinctly first-personal character. On the relevant train of thought, self-concern rests on knowing a greater amount of the same kinds of things about oneself that a subject can know about anyone; hence the explanation can only account for subjects caring *more* about themselves rather than caring about themselves *in a special way*, as the Self-Concern Intuition asserts. Moreover, the explanation invites all kinds of complications. For instance, where does the threshold between self-concern and concern for others lie, and how much can one know about another before other-concern crosses into self-concern?

Perhaps an explanation of the Self-Concern Intuition could appeal to the *kinds* of things that we know about ourselves, in virtue of our epistemic privilege. Our privileged access to our own minds means that we know a range of private and intimate details about ourselves and our inner lives which we generally do not know about others: our deepest fears, secrets, hopes, and dreams. So, the argument goes, it is our knowledge of these kinds of things about ourselves – important things, which we do not know about others, at least not with the same intimacy or immediacy – that makes us care about ourselves in a special way.

The Self-Concern Intuition still does not receive an adequate explanation. To begin with, we must take care not to overestimate the scope and importance of the kinds of facts about themselves that subjects have immediate access to. As mentioned during our discussion of the problem of other minds, not everything about our inner worlds is readily available to introspection, and

it has been forcefully argued that the more significant and complex pieces of self-knowledge may require much harder work, to the extent that they might even constitute cognitive or personal achievements (Moran 2001: xxix-xxx, Renz 2017). Moreover, appeal to such facts would still comprise too contingent and unreliable a basis for self-concern. A subject may be largely unaware of such facts about herself, or she can know as many, or even more, of these facts about others; consider, for instance, a psychotherapist who knows all the intimate details about her patients' lives but who is largely self-blind. Furthermore, on this explanation, the Self-Concern Intuition still fails to receive a first-personal explanation. If we take self-concern to be based in knowing things about oneself that one can in principle know about others, we take it to be the same kind of concern as concern for others. Consequently, this explanation of the Self-Concern Intuition bottoms out into the epistemic abundance argument: ultimately, self-concern is not due to our knowing *special* facts about ourselves, which we cannot know about others, but due to our knowing *more* of these facts about ourselves than others. Finally, and importantly, even if all these problems could be dealt with, a lot of work would have to be done to substantiate the fundamental premise of this argument, i.e. that knowledge of a specific class of facts results in caring.

We have been trying to make sense of the Self-Concern Intuition by considering the *objects* of immediate self-knowledge, particularly their kind or amount. However, what if self-concern is not about what is known, but about the way it is known? What if it is epistemic immediacy *itself* that fosters self-concern? What if we care about ourselves in a special way *because* we can access the contents of our minds directly? This is precisely the solution that is rejected by Setiya, and for good reason – it is not easy to see why knowing anything non-inferentially would result in special concern about the thing known. Setiya nevertheless attempts another solution:

Self-love is the primordial case of love at first sight. Or better, since I am available to myself not just perceptually but through immediate knowledge, in both agency and introspection, it is love at first act, or first thought. I am

presented to myself in a special and primitive way in which I am presented to no-one else: as the agent of my actions and the thinker of my thoughts. What could be more natural than to love the person who is given to me this way? (2015: 469)

After arguing that there are no specifically first-personal reasons of self-concern, Setiya proceeds to suggest that our immediate knowledge of our mental states and actions³⁵ facilitates an attitude of self-love which is justified in the same way as love for anyone else (i.e. by the fact of our humanity) but which differs from our love for anyone else because it is grounded in immediate self-knowledge. According to Setiya, this distinctive love of self gives rise to a distinctive concern for self, because love in general ‘involves a disproportionate concern for the interests of the beloved’ (469). The proposal is, then, that our immediate knowledge of ourselves gives rise to a distinctive attitude of self-love, which in turn facilitates a distinctive attitude of self-concern.

The most pressing question regarding Setiya’s proposal is the following: why would knowing our mental states and intentional actions non-inferentially make us love ourselves? What reason is there to believe that epistemic immediacy engenders self-love, and therefore self-concern? It is quite difficult to imagine the answers to these questions, and Setiya himself is notably elusive on this front. All he says is that it is ‘natural’ we should love the person who is given to us with the immediacy and intimacy with which we are given to ourselves, but this statement is neither compelling nor sufficiently substantiated. Another problem is that, in basing self-concern in love, the proposal fails to capture the first-personal character of self-concern, since love is an we can have towards any subject. Finally, it seems that this solution ultimately also attempts to locate self-concern in epistemic immediacy, albeit

³⁵ After analysing first person thought in terms of our immediate knowledge of our mental states, Setiya suggests that the latter also encompasses our capacity for immediate knowledge of our intentional actions (2015: 462, 467-468).

through the mediation of self-love, thus contradicting Setiya's earlier claim that epistemic immediacy provides no grounds for self-concern.³⁶

Let us take stock. According to the Immediate Access View, first person thought exploits our capacity for immediate knowledge of our mental states, and a nonconceptual form of self-consciousness. In this section, I have tried to determine whether our caring about ourselves in a special way could somehow be a consequence of our capacity to know our mental states directly. Several possible avenues were explored, but none succeeded in finding the basis of self-concern in immediate self-knowledge. Though this does not mean that our abundant, privileged, and intimate knowledge of our mental lives contributes nothing to the concern we take in ourselves, it is clear that the primary explanation of self-concern must be found elsewhere. The only place left to look, at least on the Immediate Access View, is nonconceptual self-awareness.

4.2.2. Pain and Pleasure

According to the Immediate Access View, a nonconceptual, nonlinguistic form of self-consciousness comprises a key feature of our conscious experience, and the most primitive form of first person cognition. Could self-concern somehow be located in, or entailed by, this form of self-consciousness? To answer this question, I will consider some of the ways in which theorists have previously tried to articulate a link between first person thought, nonconceptual self-awareness, and self-concern. I begin with Mark Textor's discussion of the work of nineteenth-century philosopher Hermann Lotze. I then move on to Lucy O'Brien's account of first person reference in *Self-Knowing Agents*, and her and Marie Guillot's discussion of self-concern in 'Self-Matters'.

In "Enjoy Your Self": Lotze on Self-Concern and Self-Consciousness' (2018), Mark Textor reconstructs, assesses, and develops Hermann Lotze's view of mental ownership, first-person thought, and the first-person concept, in

³⁶ These – and other – problems with Setiya's proposal are insightfully discussed in O'Brien and Guillot (unpublished: 18-21).

which self-concern holds a central role. The paper begins by articulating something along the lines of the Self-Concern Intuition, noting that ‘each of us takes a unique concern in him- or herself and what belongs to him- or herself’ (157). As Textor explains, an intuition of this sort is shared by Lotze, who maintains that we take a distinctive interest in what is ours, which obtains independently of our other desires, interests, and dispositions (158-159). In the discussion that follows, Textor seeks to flesh out Lotze’s understanding of the distinction between self and non-self and the way that this bears on the intuition of self-concern. The question, as Textor puts it, is: ‘What constitutes the difference between what is me or mine and what is not-me or not-mine such that what is I or mine has this unique kind of interest for me?’ (157).

According to Lotze, from the subject’s point of view, the distinction between self and non-self has its basis in an ‘immediate self-feeling’ or ‘immediate for-me-ness’. As Textor puts it,

Mental events and processes strike us as belonging to us and not to someone else or to no one at all. According to Lotze ... we can draw the self/non-self distinction because our thoughts and experiences have this immediate for-me-ness. (159)

According to Lotze, then, our conscious experience is characterised by a feeling of ‘immediate for-me-ness’, on the basis of which we are able to identify our thoughts and experiences as ‘ours’ and distinguish ourselves from the rest of the world. Moreover, and importantly, Lotze asserts that ‘all self-consciousness rests upon the foundation of direct sense of self’ (Lotze 1885: 680), suggesting that this ‘immediate self-feeling’ lies at the core of self-consciousness and first-person thought. Additionally, according to Lotze, this immediate self-feeling is both nonconceptual and conceptually and metaphysically prior to first person thought and self-consciousness (Textor 2018: 169).

In its construal of first person thought as resting on an immediate, nonconceptual ‘self-feeling’, Lotze’s account closely resembles the part of our Immediate Access View on which first person thought is said to involve a nonconceptual, nonlinguistic form of self-awareness. In fact, in taking

immediate self-feeling to provide the grounds for mental ownership, Lotze appears to construe the relevant awareness along the lines of what is today called a 'sense of ownership' or 'mineness' (Textor 2018: 159). In 4.1.3., we mentioned three ways in which the nonconceptual self-awareness purportedly involved in first person thought can be interpreted: as a pre-reflective form of self-consciousness, as a sense of ownership or 'mineness', or as an awareness of the self as the subject of experience. I endorsed the first interpretation, whereas Lotze seems to be endorsing the second one. Nonetheless, the outlines of the proposal are the same.

According to Lotze, then, first person thought rests on an immediate self-feeling, which also facilitates the subject's apprehension of the distinction between herself and the world. Where does self-concern come in, on this picture? The key to answering this question lies in Lotze's claim that immediate self-feeling, and therefore mental ownership, rests on feelings of pain and pleasure. To flesh out this claim, Lotze presents the example of a worm writhing in pain, which lacks the conceptual capacities to know or self-ascribe its own experiences, and an omniscient angel that knows everything about itself but is incapable of either pleasure or pain:

The crushed worm writhing in pain undoubtedly distinguishes its own suffering from the rest of the world, though it can understand neither its own Ego nor the nature of the external world. But the consummate intelligence of an angel, did it lack that feeling, would indeed be capable of keen insight into the hidden essence of the soul and of things, and in full light would observe the phenomena of its own self-reflection, but it would never learn why it should attach any greater value to the distinction between itself and the rest of the world than to the numerous differences between things in general that presented themselves to its notice. (Lotze 1885: 250-251; from Textor 2018: 161)

According to Lotze, then, mental ownership is grounded in feelings of pain and pleasure: it is *only* by finding an experience pleasurable or painful that a subject comes to identify it as her own, thereby grasping the distinction between herself and the world. The writhing worm, though incapable of self-ascribing its

experiences, nevertheless grasps them as *its own* simply in virtue of being pained by them. By contrast, though the angel knows everything there is to know about itself, its inability to feel pleasure or pain means that it is unable to distinguish between itself and the world, and therefore is, for itself, just one object among others (Textor 2018: 161).

We said that, according to Lotze, we take a special interest in ourselves and what belongs to us. The ensuing question was what makes an experience ‘mine’, such that I take a special interest in it. According to what we have said so far, an experience is mine if it has immediate for-me-ness, and an experience has immediate for-me-ness if it pleases or pains me. It follows that mental ownership is ultimately grounded in pain and pleasure, and since mental ownership corresponds to concern – I take a special interest in what is *mine* – concern ultimately also rests on pain and pleasure. Hence a being, like the omniscient angel, who cannot experience pain and pleasure, and cannot draw the distinction between self and not-self, cannot take a special interest in itself and its own experiences. Lotze hints at this in the passage, when he remarks that the angel does not comprehend why it should value what belongs to it any more than it values other things. By contrast, despite lacking any conceptual capacities, the writhing worm is capable of taking an interest in its experiences simply in virtue of being pained or pleased by them.

However, how are we to understand the move from being pleased or pained by an experience and thereby apprehending it as our own, to taking a special concern in it? To illuminate Lotze’s picture of mental ownership and concern, Textor helpfully recommends that we compare Lotze’s sense of ‘mineness’ with that which figures in expressions like ‘my friend’ (163). What makes someone ‘my friend’ is my loving and valuing them non-instrumentally; similarly, on Lotze’s picture, what makes an experience ‘mine’ is my taking pleasure in or being pained by it, and thereby valuing it. To take pleasure or pain in something is at the same time to evaluate it: if something pains me, it is bad, and if it pleases me, it is good (163). As Textor puts it, ‘Feelings of pain and pleasure apprehend goodness/badness; they are felt evaluations’ (163). On the

ensuing picture, affect, ownership, and concern occur all at once: to ‘own’ an experience is to value and take a special concern in it, because ownership and concern both, and together, rest on feelings of pain and pleasure. We thus arrive at an account of first person thought on which the phenomenon is constitutively, and at its very core, evaluative and concerned. This account vindicates the Self-Concern Intuition, thus showing the way – or at least one way – towards articulating an account of first person thought that fulfils the criteria put forth in this thesis.

Now, an important question: is Lotze’s explanation of self-concern plausible? Attempting an independent assessment of the account would be too extensive a project at this point, so let us rely on Textor’s assessments. According to Textor, Lotze’s conditions for mental ownership are sufficient but not necessary, because a subject can ‘own’ her experiences without being pleased or pained by them; thus the resulting account of mental ownership is incomplete (169). Nevertheless, Textor argues that Lotze’s account delivers certain crucial insights regarding our approach to first person thought and the first person concept, and specifically our tendency to focus on the phenomenon’s epistemic and referential dimensions.

First, an account of first person thought that takes its distinctive feature to be epistemic, i.e. which theorises first person thought primarily through the subject’s privileged epistemic relation to herself and her mental states, neither acknowledges nor accommodates its connection with affect and concern. This is what the example of the omniscient angel, whom Textor calls ‘Michael’, makes clear (Textor 2018: 166):

Michael’s case points us to a serious problem for the epistemic view of mental ownership: one can have non-observational knowledge about something without valuing it or being concerned about it ... If I have non-observational knowledge of x, it is still an open question whether x is of value or disvalue and hence, of interest. If I take pleasure in x, this question is closed. Taking pleasure in x is a way of evaluating x and taking an interest in it. Hence, we need to incorporate felt evaluations – enjoyment and taking displeasure – into an account of mental ownership if we want to do justice to the import of this notion.

Though Michael knows everything about himself, and does so non-observationally, he fails to take a distinctive interest in what belongs to him. This is precisely what our earlier endeavour to explain the Self-Concern Intuition in terms of epistemic immediacy revealed: epistemic privilege does not deliver concern, and theorising first person thought primarily in terms of such privilege leaves its affective and normative components unexplained. Additionally, and importantly, Lotze's discussion illuminates the importance of incorporating felt evaluations in our accounts of first person thought, especially insofar as we wish to do justice to its affectivity and normativity. As Textor makes clear, felt evaluations provide a convenient and fairly straightforward way of accounting for the concern internal to first person thought. If first person thought rests on pain and pleasure, it is fundamentally evaluative and concerned.

Secondly, according to Textor, Lotze's discussion delivers an important diagnosis regarding our tendency to focus on the reflexivity of the first-person concept (169-171). According to Lotze, the first-person concept, 'I', also rests on felt evaluations (158). However, when we reflect on first person thought, what impresses us the most is the guaranteed identity between its subject and object, and we are consequently tempted to construe the first-person concept primarily in terms of that reflexive relation and therefore in purely descriptive terms, as 'the thinker of these thoughts.' As a result, we gloss over the 'I's connection with felt evaluations and concern, failing to do justice its true content, which is much thicker than a mere reference rule could capture. This diagnosis is similar to the one offered in Chapter 2 through the category of 'Reductionist Views' – though reached via different route, the upshot is the same.

Therefore, according to Textor, the most important lesson to be learned from Lotze is that neither reflexivity nor self-knowledge are sufficient to characterise first-person thought and the first-person concept. Instead, we must endeavour to formulate accounts of the phenomenon that do justice to its evaluative dimension,³⁷ and which bring out the nature and significance of its connection with felt evaluations. In particular, the main lesson drawn by Textor,

³⁷ This lesson encapsulates the fundamental premises and the arguments of this thesis.

articulated using contemporary terminology, is that we should modify the pattern of use specifying the conceptual role of the 'I' such that it incorporates felt evaluations (173). In the paper, Textor briefly attempts this corrective project (172-174), thus making a start at providing an account of the first person concept that does justice to its affective and evaluative dimension. Therefore, even if the details of Lotze's account do not ultimately come together to provide a satisfying explanation of the Self-Concern Intuition, his insights regarding the evaluative role of pain and pleasure in first person thought pave the road for such an explanation.

We looked to Lotze for an explanation of the Self-Concern Intuition in terms of nonconceptual self-awareness. We have found a potential explanation of the Self-Concern Intuition, or at least the outlines of one, but does this explanation ultimately turn on nonconceptual self-awareness? On Lotze's view, first person thought does incorporate a nonconceptual form of self-awareness, in the shape of the immediate for-me-ness that enables the distinction between self and world. However, it is feelings of pain and pleasure that ultimately do most of the work in explaining self-concern. Does this mean that nonconceptual self-awareness is, after all, merely incidental to Lotze's suggested explanation of self-concern? I think not, because felt evaluations can themselves be understood in terms of a nonconceptual, non-propositional awareness of pain and pleasure (Textor 2018: 167), which can in turn plausibly be construed as constitutively affective and concerned.

Nevertheless, it must be pointed out that, given the more general objective of this thesis, whether Lotze's explanation – or any explanation – of self-concern ultimately ascribes such concern to non-conceptual self-awareness is not the main issue. From the outset, our goal has been to find an account of first person thought that can surrender an explanation of the Self-Concern Intuition, and the Immediate Access View was put forth as our best approximation of such an account, taking cue from Perceptual Views. However, if it turns out that a superior explanation of self-concern can be supplied by a different account, for instance an account on which concern can be traced in the

conceptual or cognitive structure of first person thought, the primary objective of this thesis will still have been achieved.

4.2.3. Agency

Another potential strategy for accounting for the Self-Concern Intuition appeals to the role that agency plays in first person thought, as found in the work of Lucy O'Brien and Marie Guillot. In *Self-Knowing Agents*, O'Brien adopts a 'two-tier' account of first person thought, on which the phenomenon is theorised partly in terms of the token-reflexive rule, and partly in terms of immediate self-knowledge and nonconceptual self-awareness. The two tiers correspond to two levels of interpretation: the first tier addresses our use of the first person concept, and the second addresses the nonconceptual forms of cognition that facilitate first person reference. O'Brien's championing of a nonconceptual self-awareness is due to her rejection of the token-reflexive rule as a sufficient analysis of first person reference on the grounds that it cannot account for its self-consciousness. It is also due to her conviction that only through an appeal to nonconceptual self-awareness could the self-consciousness of first person reference be accounted for.

The distinctiveness of O'Brien's account derives from her construal of the relevant form of awareness as an awareness of intentional action or agency. In her own words,

Our most basic awareness of ourselves is as performers of actions, mental and physical. (3)

The more primitive form of self-awareness, which it is argued can serve in an account of our capacity for first-person reference, is seen to arise from the particular awareness we have of our intentional actions. (11)

According to O'Brien, then, first person reference rests on a nonconceptual self-awareness surrendered through the subject's participation in, and immediate knowledge of, her intentional actions, both mental and physical. O'Brien calls

the relevant form of awareness *agent's awareness*, and dubs her ensuing account of first person reference 'the agency account' (77).

This focus on agency is reinforced in O'Brien and Marie Guillot's 'Self Matters' (unpublished), which examines and ultimately rejects Setiya's conclusion that self-concern does not warrant a first personal justification.³⁸ While O'Brien and Guillot agree with Setiya's conclusion that self-concern cannot be justified by the epistemic relations involved in first person thought, they also believe that those relations are not sufficient to characterise the phenomenon:

[I]t is not just that I *know* the individual I happen to be in these special ways but I am also related to that *individual's actions* in a way quite unlike the way I am related to others' actions: the individual whose actions I immediately determine in acting is the individual that is me. (5)

Therefore, in a similar spirit to *Self-Knowing Agents*, O'Brien and Guillot propose that, in addition to theorising first person thought in terms of immediate self-knowledge, our accounts of the phenomenon must also incorporate an agential element, which takes notice of subjects' immediate relation to and awareness of their actions and agency. In particular, O'Brien and Guillot suggest that first person thought should be understood in terms of a relation of 'Immediate Effect', according to which 'When I think of myself in the first-person my doing so depends on my standing to myself as the object of immediate *effect*: if in acting I immediately change something, that thing is me' (6). On this picture, 'The ability to think a first-person thought is ... based on standing in a relation to oneself as agent (*inter alia*)' (6).

How does the resulting picture accommodate self-concern? O'Brien and Guillot argue that

[W]e would, as authors of intentional change, care about the agent, and state of the agent, that is both the source and the result of those changes. If the animal

³⁸ Like Setiya, O'Brien and Guillot are interested in the justification of the Self-Concern Intuition.

that I am is required to enact, by changing, all of the things I do, how can I not care about the state of that animal? Indifference would be unintelligible, unless indifference to my actions were intelligible. But to [be] [in]different to one's agency would be to fail to be an agent at all. (9)

According to O'Brien and Guillot, then, agency is constitutively concerned. The argument offered is complex and compelling, but the upshot is that self-concern is constitutive of agency because it is necessary for the capacity to set ends and secure the means to achieve them (12, 16). According to O'Brien and Guillot, my well-being constitutes 'a *beginning* or *source* of my capacity to set ends' (12), and a lack of concern for it would annul my capacity to act at all. That is, agency would become unintelligible if detached from a non-instrumental, first-personal concern for one's own well-being.

The essence of this argument, as I understand it, is that subjects are both thinkers and agents, and, as agents, they must take a special interest in what happens to them if they are to act at all (or, at least, if they are to act as fully rational agents). Therefore, insofar as subjects figure in first person thought primarily as *agents*, i.e. insofar as the relation through which they represent themselves first-personally is that of agency or agent's awareness, their first-personal thinking manifests the distinctive self-concern that goes hand in hand with their agency. Self-concern is part and parcel of thinking of ourselves in the first person, because it is part and parcel of the agential self-relation that grounds first person thought. This approach to the Self-Concern Intuition justifies self-concern insofar as it reveals that it is a rational constraint on agency; it also explains self-concern insofar as it illuminates its operation in the subject's first-personal psychology.

4.2.4. Results

In this final part of the thesis, we have been considering whether the Self-Concern Intuition may finally receive an explanation on our proposed Immediate Access View, according to which first person thought rests on the

subject's immediate knowledge of her mental states and a nonconceptual form of self-awareness. After deciding that self-concern cannot be explained by immediate self-knowledge, we looked for an explanation of the Self-Concern Intuition in accounts of first person thought that incorporate an appeal to nonconceptual self-awareness. Here, we found two potential strategies for explaining the Self-Concern Intuition. First, on the account proposed by Hermann Lotze and expounded by Mark Textor, first person thought is essentially evaluative and concerned because it ultimately rests on feelings of pain and pleasure which comprise 'felt evaluations'. Though this account has its weaknesses, as Textor points out, it nevertheless suggests a promising strategy for vindicating the Self-Concern Intuition, both with regards to first person thought and with regards to the first person concept. Secondly, on the account proposed by Lucy O'Brien and Marie Guillot, in first person thought, subjects function primarily as agents, and are self-concerned due to self-concern comprising a constitutive component of agency.

Though it cannot be said that these accounts qualify as Immediate Access Views, or that they explain self-concern directly or exclusively in terms of nonconceptual self-awareness, they nevertheless share the following features:

- a) An acknowledgment that we must look beyond the referential dimensions of first person thought in accounting for it
- b) A rejection of immediate self-knowledge as a basis for self-concern
- c) An appeal to nonconceptual, nonlinguistic form of self-awareness

Therefore, even if the Immediate Access View is not, after all, the best account for explaining the Self-Concern Intuition, it has nevertheless played an important role in helping us articulate the fundamental features of such an account. I believe that the main achievement of this thesis consists in identifying such features, along with some strategies for explaining the Self-Concern Intuition. If that is correct, then the main objective of the thesis, namely of determining what kind of account of first person thought is best suited to accommodate the Self-Concern Intuition, has been accomplished.

Conclusion

My aim in this thesis has been to determine what account of first person thought is best suited to accommodate its practical, affective, and normative significance, and its association with a distinctive concern for self.

I began, in Chapter 1, by explaining that first person thought exhibits a characteristic practicality, since shifting to first personal forms of thought can have a marked influence on a subject's behaviour and action. For instance, when Perry shifts from 'The shopper with the torn sack is making a mess' to 'I am making a mess', this shift is accompanied by an immediate change in his behaviour. I subsequently turned to Hume's *Treatise* to illustrate that first person thought also exhibits a characteristic affectivity, as seen in its unique association with emotions like pride and shame. I argued that the affectivity of first person thought also incorporates a distinctive concern for self, manifested in the subject's 'tremorous' reaction to thoughts concerning her own future suffering. This *prima facie* association between first person thought and self-concern was referred to as the 'Self-Concern Intuition'. In Chapter 1, I argued that we must appreciate the importance of offering accounts of first person thought that make sense of its practicality, its affectivity, and its connection with self-concern, and, indeed, that our accounts must be formulated in light of this requirement. I then proceeded to explore which of the existing accounts of first person thought can fulfil these desiderata.

In Chapter 2, I proposed a category of 'Reductionist Views', according to which first person thought is adequately captured by its fundamental reference

rule, i.e. the token-reflexive rule. As I explained, although the category of Reductionist Views is largely artificial, it is also diagnostic, and intended to reflect a general preoccupation with the referential dimensions of first person thought. Here, I argued that we must shift some of our attention to considering the other features of first person thought, especially its practical and affective significance. The latter are features which a simple appeal to the token-reflexive rule has no means to explain; this means that Reductionist Views, which restrict themselves to such an appeal, fail the fundamental requirement set in this thesis. In Chapter 2, I also argued that Reductionist Views are untenable on the grounds that they cannot provide a non-circular explanation of the self-consciousness of first person reference. I suggested that the relevant explanation can be provided by the notion of a nonconceptual and nonlinguistic first person, which is nevertheless likely to be met with some skepticism, as it proposes significant revisions to the traditional philosophical understanding of thought and its relation to language.

Following my rejection of Reductionist Views, I proceeded, in Chapter 3, to consider Perceptual Views, according to which first person thought is grounded in some form of self-perception or self-awareness. I first discussed Gareth Evans' Perceptual View, according to which first person thought is grounded in the subject's awareness of her own body, which is facilitated by perceptual capacities like proprioception and kinaesthesia. As I explained, a decisive argument against Evans' account is that it fails to secure the guaranteed reference and the guaranteed reflexivity of first person thought, as it allows both for cases in which it fails to refer, and for cases in which it refers to someone other than oneself. I then considered a different version of the Perceptual View, according to which first person thought is facilitated by some sort of internal self-perception. A decisive objection against this form of the Perceptual View is that it requires the postulation of a metaphysically elusive 'self', or 'Cartesian Ego'. Having illustrated the untenability of Perceptual Views, I then pointed out that they nevertheless contain some resources for explaining cognitive

significance, namely by locating the latter in the self-relation or self-awareness that purportedly facilitates thought.

Taking cue from Perceptual Views' strategy for accounting for the cognitive significance of thought, I proceeded, in Chapter 4, to articulate the 'Immediate Access View', according to which first person thought exploits the subject's capacity for immediate knowledge of her mental states and the nonlinguistic and nonconceptual self-awareness which in Chapter 2 was appealed to as an explanation of self-consciousness. This account was articulated through reflection on the problem of other minds, and it was seen as providing a structure on which an account of first person thought can be built, rather than as a fully-fledged account in itself. After putting forth this account, I went on to determine whether it might be able to deliver an adequate explanation of the Self-Concern Intuition. After arguing that self-concern cannot be explicated in terms of the subject's immediate epistemic access to her mental states, I proceeded to ask whether it may be explicable in terms of the second component of the Immediate Access View, i.e. nonconceptual self-awareness. To answer this question, I examined two accounts of first person thought that appeal to nonconceptual self-awareness: first, Herman Lotze's account of first person thought, as discussed by Mark Textor in 'Enjoy Your Self': Lotze on Self-Concern and Self-Consciousness', and secondly, Lucy O'Brien's account in *Self-Knowing Agents* in conjunction her and Marie Guillot's account in 'Self Matters'.

My discussion of Lotze revealed that the Self-Concern Intuition may be explicable on an account of first person thought that construes such thought as resting on 'felt evaluations', i.e. feelings of pain and pleasure. Additionally, reflection on O'Brien and Guillot's account illustrated that the self-concern internal to first person thought may be explicable through an appeal to agency. Finally, I identified three features that these accounts have in common: a) an acknowledgment that we must look beyond the referential dimensions of first person thought in accounting for it, b) a rejection of epistemic privilege as an explanation of self-concern, and c) an appeal to nonconceptual self-awareness. Identification of these features served to illuminate the right approach towards

theorising first person thought, especially insofar as the requirement to do justice to its practical, affective, and normative dimensions is acknowledged. The discussion also served to bring out the kind of account of first person thought that is best suited to accommodate the Self-Concern Intuition, thus fulfilling the central objective of the thesis.

Bibliography

Anscombe, E., 1981, 'The First Person', In *Metaphysics and the Philosophy of Mind: The Collected Papers of G. E. M. Anscombe, Volume 2*, Oxford: Basil Blackwell, pp. 21-36.

Armstrong, D. M., 1984, 'Consciousness and Causality', In Armstrong, D.M, Malcolm, N., eds., *Consciousness and Causality*, Oxford: Basil Blackwell.

Barwise, J., Perry, J., 1981, 'Situations and Attitudes', *The Journal of Philosophy*, 78 (11), pp. 668-691.

Batson, D., 2010, 'Empathy-induced altruistic motivation', in Shaver, P., Mikulincer, M., eds., *Prosocial motives, emotions, and behavior: The better angels of our nature*, Washington, DC: American Psychological Association, pp. 15-34).

Batson, D., 2009, 'These Things Called Empathy: Eight Related but Distinct Phenomena', in Decety, J., Ickes, W., eds., *The Social Neuroscience of Empathy*, Cambridge, Massachusetts: MIT Press, pp. 3-15.

Bermudez, J. L., 1998, *The Paradox of Self-Consciousness*, Cambridge, Mass: MIT Press.

- Brandom, R., 2000, *Articulating Reasons: An Introduction to Inferentialism*, Cambridge Mass. and London: Harvard University Press.
- Campbell, J., 1995, *Past, Space, and Self*, London and Cambridge, Mass.: MIT Press.
- Cassam, Q., 2015, 'What Asymmetry? Knowledge of Self, Knowledge of Others, and the Inferentialist Challenge', *Synthese*, 194 (3), pp. 723-741.
- Castañeda, H. N., 2001, "'He": a Study in the Logic of Self-Consciousness', In Brook, A., DeVidi, R. C., eds., *Self-Reference and Self-Awareness*, Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 51-80.
- Castañeda, H.N., 1967, 'Indicators and Quasi-Indicators', *American Philosophical Quarterly*, 4 (2), pp. 85-100.
- Dummett, M., 1996, *Frege and Other Philosophers*, Oxford University Press.
- Dummett, M., 1973, *Frege's Philosophy of Language*, New York: Harper and Row.
- Evans, G., 1982, *The Varieties of Reference*, New York: Oxford University Press.
- Frege, G., 1918, 'Der gedanke: Eine logische untersuchung', *Beiträge zur Philosophie des deutschen Idealismus*, 1, pp. 58-77. Translated as 'Thought' in Michael Beaney, ed., 1997, *The Frege Reader*, Oxford: Blackwell, 325-345. Page references in the text are to Beaney 1997.

- Guillot, M., 2016, 'Thinking of oneself as the thinker: the concept of self and the phenomenology of intellection', *Philosophical Explorations*, 19 (2), pp. 138-160.
- Hoffman, M. L., 2008, 'Empathy and prosocial behavior', *Handbook of Emotions*, 3rd edition, pp. 440-455.
- Hume, D., 1738, *A Treatise of Human Nature*. References are to the 1978 Oxford edition of the *Treatise*, edited by L. A. Selby-Bigge and P. H. Nidditch (2nd edition).
- Johnston, M., 2010, *Surviving Death*, Princeton: Princeton University Press.
- Liberman, N., Trope, Y., Stephan, E., 2007, 'Psychological Distance', in Kruglanski, A. W., Higgins, E. T., *Social Psychology: Handbook of Basic Principles*, 2nd ed., New York and London: The Guilford Press, pp. 353-381.
- Lotze, R. H., 1885, *Microcosmus: An Essay Concerning Man and His Relation to the World*, 4th edition, transl. E. Hamilton and E. E. C. Jones, Edinburg: T. & T. Clark.
- Mach, E., 1914, *The Analysis of Sensations, and the Relation of the Physical to the Psychological*, transl. C. M. Williams, Chicago and London: The Open Court Publishing Company.
- Mellor, D. H., 1989, 'I and Now', *Proceedings of the Aristotelian Society*, 89, pp. 79-94.
- Moran, R., 2001, *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton and Oxford: Princeton University Press.

- Morgan, D., 'The Demonstrative Model of First-Person Thought', *Philosophical Studies*, 172 (7), pp. 1795-1811.
- Nozick, R., 1981, *Philosophical Explanations*, Cambridge, Mass: HUP, pp. 79-81.
- O'Brien, L., 2007, *Self-Knowing Agents*, New York: Oxford University Press.
- O'Brien, L., Guillot, M., unpublished, 'Self Matters'.
- Parfit, D., 1984, *Reasons and Persons*, Oxford: Clarendon Press.
- Peacocke, C., 2014, *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness*, Oxford: Oxford University Press.
- Peacocke, C., 2008, *Truly Understood*, New York: Oxford University Press.
- Penelhum, T., 1976, 'The Self in Hume's Philosophy', *The Southwestern Journal of Philosophy*, 7 (2), pp. 9-23.
- Penelhum, T., 1975, *Hume*, London and Basingstoke: The Macmillan Press Ltd.
- Perry, J., 2002, *Identity, Personal Identity, and the Self*, Indianapolis: Hackett Publishing.
- Perry, J., 1979, 'The Problem of the Essential Indexical', *Noûs*, 13 (1), pp. 3-21.
- Prior, A. N., 1959, 'Thank Goodness that's Over', *Philosophy*, 34 (128), pp. 12-17.

- Rawls, J., 1999. *A Theory of Justice*, Revised edition, Cambridge, Mass.: Belknap Press of Harvard University Press.
- Renz, U., 2017, 'Self-Knowledge as a Personal Achievement', *Proceedings of the Aristotelian Society*, 117 (3), pp. 253-272.
- Rogers, C. R., 1975, 'Empathic: An Unappreciated Way of Being', *The Counselling Psychologist*, 5 (2), pp. 2-10.
- Ryle, G., 1949. *The Concept of Mind*, London: Hutchinson.
- Setiya, K., 2015, 'Selfish Reasons', *Ergo*, 2 (18), pp. 445-472.
- Shoemaker, S., 1968, 'Self-Reference and Self-Awareness', *The Journal of Philosophy*, 65 (19), pp. 555-567.
- Sidgwick, H., 1874, *The Methods of Ethics*, London: Macmillan.
- Smith, J., 2020, 'Self-Consciousness', *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Zalta E. N., ed., URL = <https://plato.stanford.edu/archives/sum2020/entries/self-consciousness/> (Accessed 12 December 2020).
- Singer, P., 1997, 'The drowning child and the expanding circle: as the world shrinks, so our capacity for effective moral action grows', *New Internationalist*, 289 (28).
- Snow, N. E., 2000, 'Empathy', *American Philosophical Quarterly*, 37 (1), pp. 65-78.

- Sober, E., Wilson, D. S., 1998, *Unto Others: the Evolution and Psychology of Unselfish Behavior*, Cambridge, MA: Harvard University Press.
- Strawson, P. F., 1994, 'The First Person and Others', in Cassam, Q., ed., *Self-Knowledge*, Oxford: Oxford University Press.
- Textor, M., 2018, "‘Enjoy Your Self’: Lotze on Self-Consciousness and Self-Concern", *History of Philosophy Quarterly*, 35 (2), pp. 157-178.
- Toi, M., Batson, D., 1982, 'More evidence that empathy is a source of altruistic motivation', *Journal of Personality and Social Psychology*, 43, pp. 281–292.
- Torre, S., 2016, 'De Se Thought and Communication: An Introduction', In Torre, S., García-Carpintero, M., eds., *About Oneself: De Se Thought and Communication*, Oxford: Oxford University Press, pp. 1-20.
- Wittgenstein, L., 1958, *The Blue and Brown Books*, Oxford: Blackwell.Publishing.
- Wollheim, R., 1984, *The Thread of Life*, Cambridge: Cambridge University Press.

