Knowing Selves

Léa-Cécile Salje

A thesis presented for the degree of
Doctor of Philosophy

UCL, Department of Philosophy

I, Léa-Cécile Salje, 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.

.....

Léa-Cécile Salje

Abstract

This dissertation is about first person thought. More specifically, it's about the interface between two features of first person thought. The first is its reflexivity. First person thought is thought that is about its thinker *no matter what else might be the case* with the her at the time of thinking. This intuition of pure reflexivity is hard to shake. The second feature has to do with a special cluster of ways we have of being aware of ourselves *as* ourselves, as it is sometimes put. No further cognitive work is required, for instance, when I come to know in the normal way about the presence of hunger, anger, or blushing to put me in a position to know that *I* am hungry, angry or blushing. The deliverances of these epistemic channels come ready-formed, as it were, for uptake in a first person thought. The second feature, then, is the special tie between our first personal way of thinking of ourselves and certain forms of self-awareness.

To suggest that there is a tension between these two features would perhaps be something of an exaggeration. But by placing them side-by-side we can more readily see that there is something about holding them together than invites explanation. How, the question is, should we account for the second feature without undermining the first — where should we place these special first personal forms of self-awareness in a full account of first person thought such that they will not threaten the intuition of pure unsupported reflexivity? This dissertation develops an account of first person thinking that comfortably houses both of these features of first person thought without compromising either one.

Acknowledgements

I must begin with thanks to the AHRC and the Royal Institute of Philosophy for funding my PhD; I am extremely grateful.

More personal thanks go first to Chris Peacocke and Rory Madden, two of the supervisors on this project. Many of the issues in this dissertation interact with themes in Chris's work, and his influence over my way of thinking of them is clearly visible throughout. Rory was pivotal in guiding me to see some of the real points of interest in early drafts, and his supervisions are such that even just having him in mind as an imminent reader has often been an effective discipline on my writing. In the final year of the project I was lucky enough to spend a month working with François Recanati; I am thankful for his generous input at such an important final stage.

Enormous thanks are due to Alex Geddes and Ed Nettel, both of whom have read, commented on, listened to, discussed, added to, argued against, and only occasionally demolished much of the material that made it into the dissertation, and some that didn't. More broadly, I am grateful to the UCL graduate community for providing such a collaborative and fertile learning environment. Most of these chapters have been presented at our work in progress group, which has always been unfailingly helpful. I am thankful also for the peer support that never seems very far away at UCL; I think it's fair to say that the Žižek reading group with Tom Williams, Mog Hampson, Alex Sayegh, Ed Lamb and Tim L. Short (among others) was really more of a support group than anything else.

One way or another, Daniel Morgan has ended up reading and providing comments on almost an entire final draft. His generous feedback has been key to the final stages of the project, and has almost always led me to see new angles

in old material. I have very much enjoyed our discussions. For my title, as well as for much of my learning in the philosophy of mind over the course of my graduate studies, my thanks are due to Mike Martin.

My final and deepest thanks are to my principal supervisor Lucy O'Brien. She exemplifies an infectious way of doing philosophy that never loses contact with the original sources of curiosity and wonderment, and that leaves her students with an aspirational glimpse into how it can be done. I have never left a supervision without the bouyant feeling of fruitful ideas at my fingertips. To her I owe thanks not only for guidance through this project, but also for my continued love of our subject.

Contents

Introduction

This dissertation is about first person thought. More specifically, it's about the interface between two plausible-looking features of first person thought. A few words, first, on them.

The first such feature is that first person thought is an essentially reflexive way we have of thinking of ourselves. It is a kind of thought that is always about whoever thought it, *no matter what else might be the case* with her at the time of thinking. First person thought, that is to say, doesn't depend on the vagaries of any additional contingent empirical facts about the subject at the time of thinking, other than the fact that she is thinking the thought; it's just *always* thought that's about its thinker. This intuition of pure reflexivity seems to be at the very heart of what it seems to take to be a first person thought — even if it turns out that there are other apparent features of first person thought that might be debunked with enough philosophical legwork, the essential reflexivity of first person thought is something of a non-negotiable fixed point. It is not something easily given up. This is the first of the features of first person thought that provides the scaffolding for the set-up of this dissertation.

The second has to do with a special cluster of ways that we have of being aware of ourselves *as* ourselves, as it is sometimes put. This is to say that there are some ways of knowing about ourselves whose exercise is apt for direct response with a first person thought. No further cognitive work is required, for example, when I come to know in the normal way about the presence of hunger, anger, or blushing to put me in a position to judge that *I* am hungry, angry, or blushing; the deliverances of these epistemic channels come ready-formed, as it were, for uptake in a first person thought. These privileged first personal ways of knowing about ourselves can be bracketed out from the complete range of

ways of knowing about the things that we in fact are, where the thing known about is not necessarily given to the knower as herself. This complete range also includes such epistemic channels as seeing oneself in a backbar mirror, hearing one's voice on a pre-recorded tape, spotting one's name on a delegate list, and all our other ways of knowing about ourselves in which one must first recognise the object of knowledge as oneself before going on to form a first person judgment. This, then, is the second scaffolding feature of first person thought, that there is a distinctive cluster of forms of self-awareness that have an especially close tie to our first personal way of thinking of ourselves.

To suggest that there is a tension between these two features would perhaps be something of an exaggeration. But by placing them side-by-side we can more readily see that there is something about holding them together that invites explanation. How, the question is, should we account for the second feature without undermining the first — where should we place these special first personal forms of self-awareness in a full account of first person thought, such that they will not threaten the intuition of the pure, unsupported reflexivity of first person thought? One way of framing the basic aim of this dissertation is as an attempt to develop an account of first person thinking that comfortably houses both of these features without compromising either one. This is the undertaking of chapters 1 to 5. In the final chapter I ask how first person thought, so understood, connects to our second personal ways of thinking of one another.

This introduction provides an overview of the chapters to come. It will, I hope, be useful both as a guide to individual chapters and their interrelations, and as a plan of the overarching shape of the dissertation. I also use this opportunity to mark off some of the project's parameters.

Chapter 1 sets up an explanatory task that is taken up in the following chapters. It begins by noticing that the mode of reference determination for the first person concept and its canonical conceptual role apparently underdetermine one another. If this is right — and so long as we expect there to be *some* determinative relation between these two aspects of a concept — then it seems that we are owed an explanation. Why is the first person concept governed by *this* mode of reference determination, given its characteristic conceptual role, or

conversely, why does it have *this* conceptual role, given its mode of reference determination?

This starting challenge is formulated within a neo-Fregean framework that stays in place throughout the dissertation. Other approaches to concepts and mental content such as possible worlds or teleosemantic frameworks are not considered. The neo-Fregean framework assumes a compositional conception of thought and concepts under which concepts are the constituents of thoughts. Thoughts are the contents of psychological attitudes, individuable by considerations of cognitive significance; two thoughts differ just in case it is possible for a single thinker to take conflicting attitudes towards them at a time without thereby violating any norms of rationality. Concepts are related to their referents many-to-one, and reference is determined by the contribution that a use of the concept makes to the truth value of the thought in which it occurs. A concept can be characterised either by rules stating the fundamental condition for something to be the referent of a use of that concept, which is its mode of reference determination, or by its conceptual role, given by the canonical patterns of use that a thinker must be disposed to make of the concept in order to be counted as a competent user. I do not consider scepticism about reference, modes of reference determination, or about the existence of canonical conceptual roles.

In chapter 2 I present a positive account of first person thought designed to meet the challenge of chapter 1. The idea, at its core, is that there is more to a fully successful episode of first person thinking than merely meeting the right conditions of reference. One thing that the thinker must also do — or so I argue in chapter 2 — is to *grasp her own use* of the first person concept, or equivalently, to be in a position to know empirical facts of the form *a* is *F* about herself under the concept *I*. This is an achievement of the understanding; it is only by grasping one's own use of *I* that one can be said to undergo a comprehending episode of first person thought. To do this, I propose, the thinker must be epistemically related to herself in ways that track the referential trajectory of her thought. Even if the rule of reference for *I* cannot by itself determine that concept's canonical conceptual role, it serves to give rise to particular epistemic conditions on concept grasp. And these conditions, I show, *can* determinatively

explain that conceptual role.

The cognitive achievement of grasping one's own use of a concept is not peculiar to the first person concept. It's just that by bringing the achievement into view with respect to our uses of the first person, we position ourselves to answer the question left hanging at the end of chapter 1 — why *this* mode of reference determination, given this conceptual role, or why *this* conceptual role, given this mode of reference determination? We can now offer an explanation of the second kind: we can say that the mode of reference determination for *I* engenders certain epistemic conditions on grasping one's own use of that concept, conditions that in turn determine the canonical patterns of use that we make of the concept making up its conceptual role.

I am led by the broadness of the scope of the proposals made in chapter 2 to talk in quite general terms, as I do at times in other chapters, about *singular* thought. I mean something rather expansive by this. A singular thought, in the sense used here, is any thought that is directed at a single object, or perhaps at several single objects grouped together. It carries no implications of object-dependence, nor any metaphysical restrictions on the kinds of object it can take (e.g. concrete or abstract). It is to be contrasted with general thought (like *water is wet*) rather than plural thought (like *those steel balls are shiny*).

Even if the proposal extends beyond its particular application to the first person concept, it is not unrestricted to all singular concepts, so it is worth making explicit just how broad I take it to be. It is possible to divide ways of thinking of objects into three levels of dependence on particular environmental or contextual relations between the thinker and the object of thought. At the first level are what we might think of as context-dependent concepts proper. These are concepts whose very availability depends on the thinker's standing in the right contextual relation to her referent, including, for instance, the concepts *that*, *I*, *here*, *now*, *this*, *then*, *today*, *yesterday* and *tomorrow*. For each of these concepts there corresponds a contextual relation in which the thinker must stand to the referent in order for it to be a way of thinking about it that she has access to. Once that relation no longer holds, such concepts are no longer accessible to the thinker as a way of thinking of that object.

The second level of concepts are a grade less dependent on context than the

first. These comprise ways of thinking that still depend on the thinker bearing, or having once borne, the right kind of contextual relation to an object, but which end up being stable ways of thinking of an object that are available even when that relation no longer obtains. These include memory perceptual demonstrative concepts like *that turnstile [I climbed over yesterday]* and recognitional concepts. The availability of a memory perceptual demonstrative concept, for instance, depends on once having been perceptually related to the object in the right way, together with preservation of the perceptual memory over time. For a recognitional concept the thinker must have had enough exposure to the object, together with the workings of memory, to give rise to a stable familiarity with the object. These are, as we might put it, mixed, or combinatorial ways of thinking. I will talk about these first two levels of conceptual thought under the label *context-dependent singular concepts*, and I take it that the proposal of chapter 2 applies them both.

A third level abstracts from the contextual relations yet further. These are concepts whose availability does not depend on the thinker bearing, or having once borne, any particular contextual relation to the object. Depending on one's theory of nominal reference, perhaps nominal concepts are located at this level. I do not presume to settle that question here — what is important for our purposes is only that if there are any such concepts at this level, it is not intended that the proposal of chapter 2 extends to them.[1]

Chapter 3 is about immunity to error through misidentification. The chapter will, I hope, be of self-standing interest to theorists working with that notion. It also, however, follows closely on the coattails of chapter 2. This is because it is argued that, properly developed, the notion of immunity to error through misidentification provides a tool by which the forms of self-awareness enabling a thinker's grasp of her own use of the first person concept can be identified. I delay this task until chapters 4 and 5.

The self-standing narrative of chapter 3 converges on an explanatory challenge facing accounts of the first person concept. The challenge is to explain why it is that some first person judgments (made in certain ways) are immune

[1]The hierarchy of concepts just given is heavily influenced by Recanati's ordering of mental files in (Recanati 2012a).

to error through misidentification relative to the use of the first person concept, while others (made in others) are not. I call this *the asymmetry challenge*. I urge that the solution to the asymmetry challenge will come from a deepening of our explanatory resources. There are some explanations for the presence of immunity to error through misidentification coming from the concept's mode of reference determination, but as John Campbell has shown, these will not do in the case of the first person. The chapter's novel contribution comes in the offering of an explanation of another kind — an explanation coming not from the concept's conditions of reference, but from its conditions of understanding. Sometimes a judgment is immune to error through misidentification relative to *a* because of what it takes for *a* to refer; sometimes it is immune to error through misidentification relative to *a* because of what it takes for the thinker to grasp her own use of *a*. It is an explanation of the second kind, I suggest, that applies in the case of the first person, and so it is by appeal to such an explanation that the asymmetry challenge is resolved.

By this point in the dissertation an important role has been carved out for the forms of self-awareness that give rise to judgments with immunity to error through misidentification — that is, they enable our grasp of our own uses of the first person concept, or our comprehending episodes of first person thinking. Crucially for the above-stated aim of this dissertation, this role does not interfere with the essential reflexivity of first person thought. We still, though, do not know what these forms of self-awareness *are* for the sorts of creatures that we are. I turn to this question in the next pair of chapters, 4 and 5, that consider mental and bodily forms of self-knowledge respectively. There, I argue that the range will include, at least, our faculties of introspection and episodic memory (chapter 4), and internal bodily awareness and multimodal bodily awareness (chapter 5). The stance of these chapters is defensive; in each case I begin by demonstrating the plausibility of the claim that the faculty in question issues in judgments with immunity to error through misidentification relative to uses of *I*, and then go on to defend that claim against purported counterexamples. This piecemeal strategy is unavoidably inexhaustive, but I do not see another way of going.

Holding in place the account of first person thought developed over chap-

ters 1-5, chapter 6 turns to the question of how we should understand the special interconnectedness of first person thought, so understood, and *second* person thought. I argue, against a recent wave of support for reductionism about the second person concept, that there is such thing as distinctive second person thought, and that it cannot be reduced to thoughts of other kinds. I then undertake to show that such a non-reductionist account of second person thought can accommodate the special relation between *I*- and *you*-thought just as naturally as its reductive rivals.

A final word, before starting, on the bigger picture. This dissertation is an enquiry into the nature of first person thinking — or, at least, into the nature of first person thinking as it is for the kinds of creatures that we are. But there are much larger questions looming in the background. How should we understand the epistemology of conceptual thought? What is the relation between knowledge and representation? What sort of things must we know about an object, and in what ways, if we are to be in a position to think about it? These are big questions that could occupy a career in philosophy. This dissertation uses first person thought as a case study that provides a way in.

Chapter 1

The rule and the role for *I*

1.1 Introduction

Different accounts of first person thought begin from different starting points. One might start by noticing that first person thoughts contain uses of the first person concept, thought constituents with context-dependent reference displaying an unusual profile of referential guarantees; an intriguing combination of referential flexibility and robustness. From here the theorist of first person thought might seek to understand just how the concept works such that the reference of its uses is allowed to shift systematically between contexts, but never to lack an object. This would be an enquiry into the referential workings of the concept, or into what we might call its *mode of reference determination*: the theorist's project would be to explain what it is in virtue of which uses of the first person concept come to refer to the things that they do.[1]

A quite different point of departure is to begin by looking at the exceptionally rich and seemingly irreducible role played by first person thought in our mental economies — at the distinctive range of bases on which a first per-

[1] Although this starting point is likely to be of quite general interest, it might be of an especial interest to opponents of relativism, who hold propositional thought to involve a relation to a complete classical proposition, and who therefore deny that a proposition could be true or false only relative to a context (or subject). An example of this is the neo-Fregean movement in the 1970s and 80s to find an account of the reference determining sense of the first person that could reconcile the context-dependency of first person reference with Frege's objective and timeless realm of Thoughts. See, e.g. (Perry 1977), (Burge 1979) (Evans 1981), (Peacocke 1981), (McDowell 1984)

son thought is made directly available, and at the immediate implications for thought and action following from the entertainment of a first person thought. This second suggestion presents a natural starting point for thinkers attracted more broadly to a conceptual role semantics, under which the use that we make of a given concept or expression is treated as somehow constitutive, determining or explanatory of its meaning.

This dissertation doesn't take up either of these starting points. It begins, rather, by looking at how we should understand the relation *between* the first person concept's mode of reference determination and its canonical conceptual role, where the latter is understood as the basic deployments of the concept that a thinker must be disposed to make if she is to be counted as fully grasping the concept in question. It is clear that the two must be properly integrated. Any adequate theory of a given concept must account for the compatibility between the canonical patterns of use that we make of the concept, and facts about how its uses refer and what they refer to. The aim of this chapter is to consider a local version of this question with respect to the first person concept: what is the relation between the mode of reference determination for the first person concept — that is, facts about what the concept refers to on different occasions of use and how it refers to it — and its core conceptual role, or the canonical patterns of use that we make of it?

The aim of the next section is to bring out the general form of this challenge to say how the conceptual role for a given concept is related to its mode of reference determination. §1.3 looks at how these two aspects of the first person concept should be understood. The findings of these two earlier sections are brought together in §1.4, where the question is posed directly about the relation between these two aspects in the case of the first person concept. A statement of the starting explanatory task, to which the next few chapters offer a response, is given in §1.5.

1.2 The general form of the challenge

Our ability to use singular concepts to think about the world we inhabit and the things it contains is of central importance to the kinds of creatures we are.

It is partly by virtue of a capacity to think about individual objects, persons, times and places that we are able to navigate our spatio-temporal environments, make plans and promises, maintain meaningful relationships, remember where we've been and what we did there.[2] Despite this centrality to our lives and activities, however, there is something deeply mystifying about how it is that concepts allow us to think about things in the world. How could these patterns of structured mental representations in the domain of thought put us in touch with the public realm of concrete objects? And what determines these thought-world relations? What facts fix which representation refers to which worldly object?

This last is a question about reference determination, about how our token uses of concepts refer to the things that they do. There is no universally accepted approach to this question in general, and even within a single approach there is no reason to think that there will be a single answer for all different singular concept kinds — reference comes in many varieties. The correct story about what it is in virtue of which my use of a nominal concept refers to the thing that it does is likely to look very different to an account of the facts that determine the reference of my use of a perceptual demonstrative, or a second person concept.[3]

This dissertation will be developed within the parameters of a broadly Fregean framework. In it, the content of a thought is given by a suitable statement of the conditions in which it would be true, with the reference of its constituent concepts given by a suitable statement of the conditions determining the concept's contribution to those truth conditions. We can do this by the specification, for each concept, of a rule stating the determinative conditions on the contribution of any given use of the concept to the thought containing it.[4]

These rules will look very different for different kinds of singular concepts. One natural dividing line falls between context-insensitive concepts whose

[2] 'Partly' because it is plausible that our capacity for mental states with non-conceptual content is also partly responsible for these kinds of capacities.

[3] If there is such a thing; see Chapter 6.

[4] I take this to be a fairly standard way of setting things out in the Fregean tradition, but am influenced most heavily in my characterisation of the framework by Peacocke's more recent work, and in particular his theory of concepts given in *Truly Understood* (2008) and elaborated in *The Mirror of the World* (2014).

reference remains stable across changes in context of thought, and context-sensitive concepts like *you*, *here* and *that*. An example of the former kind is the nominal concept *Mount Everest*, whose reference remains constant across different contexts of use.[5] How we should understand the determination of nominal reference is far from a settled question, but under a thought-theoretic analogue of the received view passed down from Kripke it is because my use of the concept is causally related in the right way through communicative chains to an initial baptism of the object with that name in my linguistic community.[6] Supposing this to be the correct account of nominal reference, it will be facts of this kind that will enter into the rule of reference determination, or the rule encoding the impact that a use of the concept *Mount Everest* will have on the truth conditions of a thought containing it: it is the object to which I am causally related in this way that is contributed by this concept to the determination of the truth value of any thought containing it.

Context-sensitive concepts can themselves be separated into those whose determination of reference depends on a special kind of fact about the token use of the concept itself, and those that don't. Those of the second kind correspond in the conceptual realm to the linguistic grammatical category of *true demonstratives* (Kaplan), or *discretionary indexicals* (Perry), paradigmatically manifested in episodes of thinking in which one thinks of something as *this*, or *that*. *That* is a context-sensitive concept, in that its reference shifts between different contexts of use. This context-sensitivity is reflected in the rule of reference determination by mention of a suitable perceptual, or perhaps perceptual-attentional relation that must hold between the thinker and an object if it is to be counted as the referent of her thought — which object is contributed to the truth value of the thought containing a use of *that* will vary relative to changes in the object to which the thinker bears that contextual relation.

Context-sensitive concepts of the first kind, by contrast, correspond to the category of *pure indexicals* (Kaplan), or *automatic indexicals* (Perry), and include such concepts as *I*, *now* and *today*. Uses of these indexical concepts similarly

[5] The 'context-dependent' and 'context-independent' concepts discussed here correspond to levels 1 and 3 from the introduction, see pp.16-17.

[6] See Sainsbury and Tye for a direct application of this model to the mental domain; (Sainsbury and Tye 2012).

vary with context, but in a more systematic way than the demonstrative concepts. Rather than depend on contingent contextual relations holding between a thinker and her environment, the variation of reference across contexts for this group of concepts is itself a stable feature of their uses. For a use of *now*, for instance, the contribution to the truth conditions of a thought containing it is given by the rule that its uses always refer to the time of use; for *today*, the day of use; *I*, its subject. The rules encoding the determination of reference for these concepts each make mention of features of the token use of the concept itself — about the time, the day or the subject of the token's occurrence — and as such have also been discussed by Reichenbach and others following him as *token reflexive* concepts.[7]

The determination of reference is a fundamental aspect of the Frege-inspired conception of senses, or concepts. Frege's commitments to the principle that sense determines reference, and to the idea that such senses are to be spelled out in truth-conditional terms, are characteristically adopted as core features of neo-Fregean accounts. It is not, however, the only important aspect of such accounts. Senses, or concepts, are also answerable to considerations of cognitive significance, and the thoughts containing them also serve in the role of the contents of psychological attitudes. There has been not a little scepticism since Frege, and especially in the last thirty years or so, about whether a single sort of thing could perform all of the roles originally allocated to senses, which has typically led to a fractioning of labour between different semantic elements of thought.[8] Whether or not this scepticism is justified, what it brings out is that a full account of a given concept must go beyond a specification of the concept's mode of reference determination. It must also say something about the concept's currency in the thinker's cognitive economy.

For Frege and his closest followers this is to be spelled out in terms of what Gareth Evans called the intuitive criterion of difference, the principle of thought individuation according to which two thoughts differ iff it is possible for a thinker to take conflicting attitudes towards them at a time without thereby

[7](Reichenbach 1947)
[8]See, e.g. (Perry 1977), (Burge 1979), (Bell 1979), (Heck 2002)

violating any norms of rationality.[9] If thoughts, composed of their constituent concepts, are to serve as the contents of psychological attitudes, however, then we might want to look beyond this to the broader role taken on by thoughts containing the concept in question within a thinker's psychological activities. In particular, a fully adequate account of a given concept will look to what we might think of as the canonically justified moves in thought and action associated with the use of a given concept — what are the canonical grounds on which a use of the concept becomes immediately available, and what the canonical implications of its deployment? An answer to this question constitutes an account of the core or canonical *conceptual role* of the concept.[10]

A satisfying account of a given singular concept, then, will say something both about the way it functions as a mechanism of singular reference (its mode of reference determination) and about the kinds of movement in thought canonically associated with its use (its conceptual role). It is, moreover, a constraint on the adequacy of such an account that these two aspects are construed as compatible with one another. If our account of the ways in which we use a given concept turned out to be at odds with the ways in which we would *expect* to use it based on what we take its uses to refer to, we would have a clear indication of having gone wrong somewhere.

This integrative requirement that applies to accounts of the conceptual role and the mode of reference determination for a given singular concept can be illustrated by analogy with the logical connectives. Take the case of conjunction. There are at least two ways to approach questions about how we should understand this operator. One way to go would be to look at its truth table, that provides an exhaustive listing of the various contributions that its use will make to a proposition or sentence containing it. We may take this to be analo-

[9](Evans 1982, pp.18-19); the intuitive criterion of difference is originally formulated in terms of the thoughts expressed by two sentences.

[10]'Canonical conceptual role' is to be understood here in terms of the basic deployments of a concept that a thinker must be disposed to make in order to be counted as a competent user of the concept. Is this conceptual role normative or purely descriptive? Indeed, does it even make sense to suppose that there can be norms associated with such non-rationally motivated movements in thought? If it does, could it be that these norms are not incorporated into the conceptual role itself, but merely supervene on it? These are important and daunting questions, but not ones that I need take a stance on here. I will talk as if the conceptual role is purely descriptive of the normal patterns of use that we make of a concept.

gous to an account of the mode of reference determination for a singular concept. Another way to go would be to look not at its truth table, but to the rules specifying the kinds of uses we can legitimately make of it. We might begin, that is to say, by getting its introduction and elimination rules clearly before us. This corresponds analogously to the conceptual role of a singular concept.

Now, it is evident that the truth table and the inference rules for a connective like conjunction cannot be treated independently of one another, but just how we should understand their interdependence is a further question. Most modestly put, it might be construed as a negative claim — it is a constraint on a satisfying account of conjunction that there is no incompatibility between its truth table and inference rules; the ways in which we in fact use the connective must not conflict with the ways we should expect to use it, given its truth table. This seems right. Any incompatibility between the two must plainly lead to scepticism about whether we have succeeded in getting a handle on a single unified phenomenon. This negative condition, however, seems not yet to fully capture the intuition of the interdependence, which surely goes beyond a mere requirement of consistency. It is not just that the inference rules must be *compatible* with the truth table. It is extremely plausible, rather, that there is some *explanatory* or *determining* relation between them.

Which way round this relation is to go is a matter for substantive philosophical enquiry. A likely looking picture in the case of logical operators, however, is that *these* are the inference rules for conjunction because *this* is its truth table. The reason we use the operator in the ways that we do, that is to say, is explained by the contribution it makes to the truth value of propositions or sentences containing it.

This is the suggestion made by John Campbell, who offers the following story to illustrate the point:

> Suppose that you are teaching a class elementary logic from scratch. One way to begin is simply to spend the first couple of weeks drilling the class in the inference rules for the propositional connectives, without any concern at all for their intuitive meanings. Once the class has mastered the formal manipulations associated with the terms, you introduce the truth-tables for those terms. At this point, there may be a certain sense of illumi-

> nation, as the intended meanings of those signs are revealed. At this point, you can explain to the class why the rules of inference you have introduced are not simply arbitrary. You can point out that, given the truth-tables, the rules of inference you have introduced are the weakest possible introduction rules that guarantee the truth of a statement containing the constant, given the truth of the undischarged premises; and the elimination rules are the strongest possible that guarantee the truth of the conclusions, given the truth of a premise containing the constant. (Campbell 2012, p. 2)

Once supplied with the truth tables, the idea is, the students will gain an understanding of *why* the inference rules are the rules that they are, and their application of the rules will become causally sensitive to their knowledge of the truth tables. They will be in a position to (re)derive the inference rules from the truth tables, or will be ready to make suitable adjustments to the rules in response to stipulated changes to the truth tables. Even in a case in which the patterns of use are learned first, then, it still seems as if they are explained by their truth tables, and not the other way around. That we use the logical connectives in the way that we do seems naturally explained by the fact that they mean what they mean according to their truth tables.

This case of conjunction brings out the general form of the question that I want to explore with respect to the first person concept. It is clear that the mode of reference determination for a given concept must be compatible with the canonical patterns of use that we make of it; the correct characterisation of either one of these aspects of the concept mustn't rule out the correct characterisation of the other. But the case of conjunction shows that what we're really after is something more than mere compatibility. After all, it is easy enough to see that the mode of reference determination for a given concept *a* may be *compatible* with the conceptual role for a distinct concept *b*, insofar as neither constrains the other. What we're after when we ask after the relation between the mode of reference determination and the conceptual role for one and the same concept is presumably something stronger — and more interesting — than this.

Suppose we agree that the relation between the mode of reference determination for a given singular concept and its conceptual role outstrips a negative compatibility constraint. How, then, should we positively understand the relation? What secures alignment between the way in which a concept functions to

refer in different contexts of use, and the canonical movements in thought and action associated with its use? The two most obvious suggestions are either that its conceptual role constrains what we can say about its mode of reference determination, or that things are the other way around. A third solution submits a holistic explanation; perhaps the concept's conceptual role and its mode of reference determination should be understood as mutually explanatory or determining of each other. I will consider each of these directions of explanation in relation to the first person concept in §1.4.

Before turning to the first person concept, though, I want to make a quick comment on why it is that the general question set out in this section about the relation between the conceptual role and the mode of reference determination for singular concepts is of particular interest when applied to the case of *I*, and perhaps to the other token reflexive concepts too. These cases are of special interest because of the staggering contrast between the austerity of the rule of reference determination, and the richness and complexity of its conceptual role. As Christopher Peacocke writes,

> [T]he first person may at first glance appear to be a salient counter-example to the thesis that fundamental rules of reference for a concept can contribute essentially to the explanation of normative and reason-involving phenomena involving the concept. Articulating the phenomena distinctive of the first person has drawn forth some of the most striking contributions from the greatest philosophers [...]. It must seem stretching credulity to suggest that this range of phenomena can be explained merely by drawing upon the simple reference rule that a use of *I* in thought refers to the thinker, the producer of the thinking, together with auxilary hypotheses.(Peacocke 2008, p.77)

Peacocke's interest here is somewhat broader than just the conceptual role of *I*, but it is likewise quite remarkable to think that such a simple rule of reference determination could be enough to determine the intricate patterns of use that we make of the first person concept. And as we will see, it also seems as if that rich and complex conceptual role underdetermines the rule of reference determination for *I*. There seems, then, to be a special difficulty in the case of the first person concept (and others like it) in giving a satisfying account of the

relation between the mode of reference determination and its conceptual role. The aim of §§1.4-1.5 is to bring out this difficulty.

First, though, I will offer a clearer statement of what I take the mode of reference determination and the conceptual role of the first person concept to be.

1.3 The first person concept

1.3.1 Reference determination

The rule of reference determination for the first person concept is a rule of systematic subject-reflexive reference. We can put the rule as follows.

> *Reference determination for I:* Uses of the first person concept refer to the producer, or thinker, of the thought containing it.

This rule specifies the relation of *being the thinker*, or *producer*, of the thought containing a token use of the first person concept as the constitutive relation in which one must stand to the token in order to be counted as its referent.

In ordinary circumstances we understand well enough when the condition of being a thought's thinker or producer has been met, but neither formulation is without its complications. Depending on how one understands the notion of being a producer, there are all sorts of people who might plausibly be said to produce any one of my thoughts, including those containing uses of the first person concept. Suppose you pop your head round my door to remind me that the meeting started five minutes ago. *I'm late!*, I think. There is a sense here in which you are the producer of my thought (and my first person thought at that) which is caused in rational response to what you have just told me. But this, of course, is not what we mean when we say that uses of the first person concept refer to their producers. Perhaps cases of this kind could be ruled out by a suitable restriction on the notion of production to an immediate, non-rational causal process. Counterexamples to this suggestion, however, need only be made a little more colourful: we can imagine a brilliant but deranged neuroscientist who has set things up in such a way that she has the power to remotely

manipulate my thought output. When she causes me — immediately, without engaging my rational faculties — to think a first person thought, she is its producer in this more restricted sense. This again gives us the wrong result; we do not want to say that my uses of the first person concept refer to her.

Formulations appealing to the relation of *being a thought's thinker* avoids these problems, but comes with difficulties of its own. Prereflectively, we might think of being a thought's thinker as being the person in whose stream of consciousness the thought occurs. The well-documented schizophrenic delusion of thought insertion, however, reveals that things are not quite so simple.[11] Sufferers of the delusion experience some thoughts as being generated by an external locus of agency (a person, an inanimate object, an institution) and implanted into their directly introspectible mental streams. It is, in fact, unclear whether or not these inserted thoughts ever contain uses of the first person concept — more typically, they come in the form of imperatival orders, or second- or third-personal critical narratives. The most natural reference ascription for an inserted first person thought, if such thoughts there could be, however, would presumably not have it as referring to the person in whose stream of consciousness the thought occurs, but to the imagined source or agent of the thought. Some further precisification, then, would be needed to capture the notion of *being a thought's thinker* as it is drawn on in the rule of first personal reference determination if we are to avoid getting the wrong results in such pathological cases.

There will surely be ways of tightening the notion of being a thought's thinker, or of being its producer, such that these sorts of difficulties can be avoided. Or perhaps these issues might be circumvented altogether by appeal to a different but adjacent notion, like that of being the thought's *agent* (though there will perhaps again be complicating questions for this suggestion about what to do with passively occurring thoughts). I will not pursue these questions here, but will rather limit myself to the intuitive understanding of what it takes to be a thought's thinker or producer with which we started.

The above rule of reference determination presents the first person concept as a device of systematic subject-reflexive reference. This picture is, I think it

[11] See §4.1.1 for further discussion of the phenomenon of thought insertion.

is fair to say, widely accepted, and I will not argue for it further. It is, however, worth noting an alternative view. This is a view held most notably by Gareth Evans, but that is otherwise less often self-attributed than offered in the hands of an imagined opponent. On this view the mode of reference determination of our first person thoughts follows the model of demonstrative reference: uses of the first person concept are determined to refer to whichever person it is to which the user of the concept bears the relevant perceptual-epistemic relations. So long as the relevant relations are self-bearing, the referential predictions of this model will converge with those of the rule given above, and so will explain the normal patterns of reference ascription just as well. What, then, favours what we might call the self-reference rule model over this demonstrative model?

The first thing to say appeals to a consideration of priority. Although the predictions of the demonstrative model might seem no less correct than those of the self-reference rule model, it's hard to ignore the hunch that they are only correct insofar as they agree with the predictions of that model. As Campbell puts it:

> Alternative accounts of how the reference of 'I' is fixed are satisfactory only insofar as they agree with the determination of the token-reflexive rule. Insofar as they disagree with the determination of the token-reflexive rule, alternative accounts of reference-fixing are invariably wrong and the token-reflexive rule invariably gets it right. (Campbell, 2012, p.6)[12]

If the verdicts of the two models of reference determination were ever to come apart, the thought is, then it is surely the rule of subject-reflexive reference and not the trajectory of the perceptual-epistemic channels that would secure the pattern of reference ascription we recognise as characteristic of first person thought.

Speaking more generally though, problems with the demonstrative model are well documented.[13] At the root of many of these problems is the model's rendering of first person reference as hostage to the unstable movements of

[12]The same point is made in (Campbell 1994, p.125)

[13]See, for instance, (Campbell 1994, Chaps 3 and 4), (Anscombe 1997), (O'Brien 1995).

worldly information channels. The vulnerability of putting things this way shows up in at least two sorts of case. The first involves a lack of incoming information from the referent; the most well-known example is in Anscombe:

> 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!" (Anscombe 1997, p. 153)

Assuming this outburst to be the expression of an underlying *I*-thought, the demonstrative model theorist must find a way to explain how it could be that the sensorily deprived subject's capacity for first person thought survives the suspension of all incoming sensory information from herself.

The second kind of case involves not the suspension of self-directed information channels, but their diversion to another subject. It does not seem especially difficult, at least *prima facie*, to conceive of a case in which a subject is wired up to another's body such that she receives information about it in just the same way that she normally receives information from her own body 'from the inside'.[14] The demonstrative model theorist faces a challenge to explain our reluctance to say that the reference of our first person thoughts in such a set-up would accompany the shift in the informational source.

The demonstrative model theorist, of course, is aware of such cases; much of their work lies in finding ways to counteract them, and they are certainly not without things to say.[15] There is, however, a yet even broader comment to make about why we still should not be tempted by the model. That is, that one way to see the model is as an uncompromising backing of one of the two intuitions with which this dissertation started, leaving the other somewhat unprotected. In the introduction (pp.13-14) I set out something like a starting puzzle

[14] I discuss these cases in more detail in §5.1.1, where I argue in that this apparent conceivability is in fact illusory.

[15] Evans, for instance, deals with information-suspension cases by weakening the information requirement to a *dispositional* requirement (Evans 1982, p. 216), and more recently Daniel Morgan has responded to information-redirection cases using a knowledge maximisation principle (Morgan 2015a). Another way to go would be to restrict the relevant perceptual-epistemic relations to introspection, a move resulting in a view that does not seem vulnerable to the same sorts of counterexamples.

about first person thought: how are we to bring together on the one hand the intuition of pure reflexivity, and on the other, the intuition that our first person way of thinking is deeply enmeshed with certain forms of self-awareness. The demonstrative model takes the second of these intuitions in earnest. First person thought *just is* systematic reference to the source of that self-awareness. It does so, however, at the risk of leaving unsubstantiated the first of the intuitions. Even if the results of the two models happen to coincide, it is not *because* first person thought has a purely reflexive form that it always ends up referring to its thinker. But the intuition of pure indexicality is hard to shake. The first person pronoun is often construed as a paradigmatic example of a Kaplanian pure indexical, which is to say that its reference seems to be determined independently of any special intentions or actions on the part of the speaker. Determination of first person reference in language is, as Perry says, automatic. Likewise, at the level of thought, the idea is extremely powerful that any use of the first person concept will succeed in referring to its thinker, no matter what else might be the case with the thinking subject. The demonstrative model of first person thought does violence to this picture of the first person concept as a mechanism of freestanding or unsupported pure indexicality.

1.3.2 Conceptual role

The conceptual role, or normal patterns of use that we make of a concept can be filled out in terms of the canonical inputs and outputs associated with its use. On the input side there are, as Imogen Dickie puts it, the *basic deployments* of a concept, or the deployments that a thinker must be disposed to make if she is to be counted as fully grasping the concept in question. These basic deployments, for Dickie, are those that are canonically justifiable, which is to say that 'to count as grasping the concept you must be disposed not just to make the deployment, but to make it on the grounds of a specific ('canonical') justification.' (Dickie 2011b, p.1) The article from which this quote is taken is concerned to show that perceptual demonstrative concepts have such canonically justifiable basic deployments. Crudely put, if a thinker is to be counted as grasping a perceptual demonstrative concept, then she must be disposed to deploy it on the grounds

of perceptual attention to an object. The task now at hand is to establish the parallel basic deployments of the first person concept.

Following Dickie, this task can be refined by the question, what are the grounds on the basis of which a conceptually competent thinker must be disposed to make use of the first person concept? What, as we might put it, are the forms of awareness that are canonically apt for immediate response with a first person thought by a competent thinker?

Taken as a request for a description of our ordinary patterns of use of the concept, an answer to these questions will emerge not from philosophical argumentation, but from reflection on the ways in which we actually deploy the first person concept in thought as users of the concept. What, then, does such reflection deliver as the forms of self-awareness on the basis of which a first person thought becomes immediately available, where there is no seeming gap between possession of the grounds and the disposition to make use of the first person concept? The range is surely vast, but I take a few core examples to include the way in which one knows of one's own blushing 'from the inside', the detection in the normal way of hunger, thirst, or fatigue, the introspective realisation that one has been daydreaming, proprioceptive apprehension of crossed legs, the distinctive awareness we each have of our mental and bodily actions, kinaesthetic perception of the movement of one's typing fingers, the sensation of a headache coming on, directly experienced pangs of guilt, and so on. The list is, of course, near-endless.

Notice that it won't help to explain why the input side to the conceptual role for *I* has the composition that it does — why it is *these* forms of self-awareness that enter into the conceptual role, and not others — to say that these forms of self-awareness already have first personal content. Such a suggestion is less than explanatory in at least two respects. First, it requires elaboration before it could be used in any such explanation. What does it really mean to say that visual, or proprioceptive input comes ready formed with first person content? As Campbell writes, 'there is no term literally showing up in one's perception at all; it is not as if vision is subtitled' (Campbell 1994, p. 118)[16] And second,

[16]Campbell's point is made for the first person pronoun, of course, but the same call to explain just what the proposal comes to applies just as much for the first person concept as it does

even if we could make sense of the proposal, it does nothing more than push the explanandum back a level. Why do only *these* forms of awareness have first personal content, and not others?

The second aspect of the conceptual role is the output side. What are the mental and bodily actions that are apt to follow immediately from a deployment of the first person concept in thought by a competent thinker? Again, it will be illustrative here to consider first the case of the perceptual demonstrative. Which actions follow in this way from a use of a perceptual demonstrative concept are infomed by the fact that the basic deployments of the concept are those made on the basis of a perceptual-attentional relation holding between thinker and object — the actions that follow are those that are canonically motivated by the holding of such a relation. Typical examples might be reaching out to grasp the object, or the redirection of one's trajectory in order to avoid bumping into it, or a shifting of one's position so as to better appreciate its aesthetics; actions all that are made appropriate by the relation in which the thinker stands to the object.

What about the case of the first person concept — what sorts of actions do uses of *I* render directly appropriate? The actions of interest here are those made familiar by Perry and others following him. The standard way of bringing them out is to describe a case in which a subject already thinks something of somebody, but comes to realise that — as the subject would express it — that somebody is *me*. A classic example comes from Ernst Mach, and is discussed by Perry:

> Not long ago after a trying railway journey by night, when I was very tired, I got into an omnibus, just as another man appeared at the other end. 'What a shabby pedagogue that is, that has just entered,' thought I. It was myself: opposite me hung a large mirror. The physiognomy of my class, accordingly, was better known to me than my own. ((Mach 1914), cited in (Perry 1990, p.1))

There is a clear shift in the propriety of certain actions following Mach's recognition of the man he sees as himself, and it is not difficult to know how to go

for language.

about filling in the details — perhaps he will brush down his trousers, check his posture, resolve to get a haircut, delight in his sense of identification with his social class, or (as Perry imagines the case) pick lint off his vest, actions all that are only made available by his realisation that the person he sees is himself. These are actions that become available immediately following the use of a first person concept in thought.

Such cases suffice to give us an intuitive handle on the sorts of thoughts and actions that become newly appropriate following the pivotal realisation in these cases, amounting to a basic grip on the output side of the conceptual role of the first person concept. With this in place, we can characterise the conceptual role for the first person concept as follows.

> *Conceptual role for I:* The first person concept's conceptual role is constituted of the canonical forms of awareness apt for immediate response with a use of the first person concept, and the canonical forms of action apt to immediately follow from such a use.

For reasons of scope, I will focus in this dissertation almost entirely on the input side of the conceptual role.

A note on the formulation just given. This way of putting it focusses explicitly on *immediate* inputs and outputs canonically associated with uses of the first person concept. But, of course, deployment of the first person concept is not limited to these moves. I need not be responding to an informational input of the kind included in the above range in order to think of myself first personally; I can perfectly well think *I*-thoughts in contexts other than those in which I am responding to bodily sensations 'from the inside', or making an introspectively formed judgment, or one on the basis of an episodic memory. As Evans makes the point, 'I can grasp the thought that I was breast-fed, for example, or that I was unhappy on my first birthday, or that I tossed and turned in my sleep last night, or that shall be dragged unconscious through the streets of Chicago, or that I shall die' (Evans 1982, pp.208-9). The contents of these thoughts tells us that they were all formed in ways other than those included in the privileged range described on p.35. But they are, of course, thoughts all no less first personal than the proprioception-based thought that my legs are crossed, or the

introspective judgment that I see a canary — indeed, for Evans such thoughts are of paramount importance to an account of first person reference; a creature that lacked the capacity to think thoughts of this kind could not think of themselves, as themselves, as elements of the objective order.[17]

It is, of course, right that I can think of myself first personally as the protagonist of these situations and others like them. There is, however, something mediated, or indirect, about these uses of the first person concept. Unlike the earlier cases, I cannot respond *immediately* to a photograph of an unhappy one-year-old with a use of the first person concept — I must first make an identificatory judgment that that child is *me*, the person I think about in the way constrained by the canonical moves described above. I must, that is to say, draw on additional information about the situation. Identifications, however, are cheap; they tell us very little about how the concepts on either side of the identification are to be used in the first place. In terms of characterising the conceptual role of the first person concept, this renders these cases of secondary importance. Even if there is a sense in which we can say that it is part of the normal patterns of use of the concept that it can be involved in identifications of this kind, it is the direct and unmediated uses of the concept that serve to directly delineate its conceptual role.[18]

We now have before us two things. First, an explanatory question: what is the relation between the mode of reference determination and the conceptual role for a given singular concept? And second, accounts of the two elements involved in this question — the mode of reference determination and conceptual role — in the case of the first person concept. The next section brings these materials together. How, it asks, should we understand this relation in the case of the first person?

[17](Evans 1982, p.210)

[18]One way of understanding these identification-involving uses of the first person concept is in parallel Peacocke's *cantilevering* treatment for observational and sensational concepts, under which applications of the concept in response to non-core bases should be understood as involving tacit knowledge, on the part of the concept-user, of a sameness relation holding between the object of the core and of the non-core uses. See (Peacocke 2008, p.221)

1.4 The reference rule and the conceptual role of *I*

Suppose we accept, given the considerations of §1.2, that the relation between the mode of reference determination for a given concept and its conceptual role exceeds a mere non-incompatibility constraint. There are then at least three ways of understanding the relation between them. The conceptual role might determinatively explain the mode of reference determination; the mode of reference determination might determinatively explain the conceptual role; or the two might be mutually determinatively explanatory of each other. I will consider each of these options in turn with regards to the first person concept.[19]

1.4.1 Role determines rule

The first way round of putting things is that the conceptual role for the first person concept — the normal patterns of use that we make of it, comprised of its canonical inputs and outputs — is in some sense explanatorily prior to its mode of reference determination by the rule of subject-reflexive reference. The general form of this view is one that is shared between Peacocke's earlier theory of concepts as outlined in *A Study of Concepts* (1992) and Campbell's most recent writings on the first person.[20] I will use Peacocke's early writings as a representative example of the view.

For early Peacocke, this way of organising things is not special to the first person concept. His aim in *A Study of Concepts*, rather, is to provide and defend a general theory of concepts on which the nature of any given concept is exhausted by its possession conditions — a view broadly motivated by the

[19]For each of these proposed directions of explanation, the question is not whether we can *fully* derive the one from the other, but only whether the one aspect features centrally in a derivation of the other. For one thing, it is likely that the full explanation, in any of these directions, will need to draw on empirical considerations about the forms of awareness mentioned in the conceptual role. Even if it is possible to derive facts about the general form of the conceptual role from the rule of reference determination, for example, the forms of awareness featuring in the resulting conceptual role might look very different for a creature who was set up with radically different ways of knowing about herself to the forms of awareness available to creatures like us. In that case, a full explanation of the conceptual role would need to appeal both to the rule of reference determination, and to particular features of these forms of awareness.

[20](Campbell 2012)

idea that anything that slices more finely than possession conditions can only make a difference to the nature of a concept that makes no difference for us. Specification of the possession conditions for a concept, for Peacocke, involves a specification of the transitions in thought that a competent user of the concept must find primitively compelling. These possession conditions, then, make up at least part of what we have been calling a concept's conceptual role.

For Peacocke, these possession conditions must determine — together with the world — a semantic value for uses of the concept, or a way of determining the concept's reference. He takes this requirement to follow from combining the claim that concepts are individuated by possession conditions together with the Fregean commitment to the idea that concepts (with the world) determine reference. This is the so-called *requirement for a determination theory*: the need to provide 'for each concept a theory of how the semantic value of the concept is determined from its possession conditions.' (Peacocke 1992, p.17) As it turns out, the determination theory can be stated in general terms:

> The determination theory for a given concept (together with the world in empirical cases) assigns semantic values in such a way that the belief-forming practices mentioned in the concept's possession conditions are correct. That is, in the case of belief formation, the practices result in true beliefs, and in the case of principles of inference, they result in truth-preserving inferences, when semantic values are assigned in accordance with the determination theory. (Peacocke 1992, p.19)

Rather more losely put, the assigned determination of semantic value is the one that makes best sense of, or vindicates, the primitively compelling transitions mentioned in the possession conditions.

The Peacocke of *A Study of Concepts*, then, takes the conceptual role of any given concept (or, at least, the input side of it) to determine the way in which its reference is determined. What should we make of this picture applied to the first person concept?

The first thing to say is that there are a number of problems with this view construed as a theory of concepts more generally, some of which are raised by Peacocke himself in a later study guide on the philosophy of language (Peacocke 1998). A first set of problems concerns attempts to render the conceptual

role for a given concept *autonomous*, which would be to say that the correctness of a given conceptual role is not subject to referential constraints. One of these is what he calls the *problem of spurious concepts*: the objection that such a view will force us to recognise certain concepts as meaningful even though their patterns of use do not converge on any single unified referent. If the conceptual role is determined independently of considerations of reference, there seems to be no principled reason for which we could rule out the possibility of such spurious concepts.[21]

A second problem with counting the conceptual role of a concept as autonomous is that it seems to entail that the only propositions containing a use of the concept that can be evaluated as true or false will be those that can be derived from the transitions mentioned in the concept's conceptual role — on this view of things, as he puts it, 'there should be no place for a correct (true) arithmetical [proposition] not establishable as such on the basis of materials in the canonical conceptual roles of its constituents.' (Peacocke 1998, pp. 100-101). Gödel's first incompleteness theorem, however, reveals that this cannot be the case; 'Even in the domain of arithmetic and higher-order logic, not every sentence which is true can be derived solely from the principles acceptance of which is required for understanding of the expressions in the sentence.' (Peacocke 1998, p.100). This is because Gödel's incompleteness theorem shows that for any effectively generated arithmetic theory, there will be true arithmetical statements that cannot be proved within the theory.

Both of these problems pertain to the question of whether the correctness of the conceptual role can be established independently of considerations of reference. A more lose-footed way of putting what should concern us here is, if patterns of use are unconstrained by reference, what could possibly ensure their convergence on single and unified referents? And how can we settle on the truth value of propositions containing the concept, if it is left undecided by the transitions mentioned in the conceptual role in terms of which the mean-

[21] Although this is a problem for conceptual-role-first views in general, it is one that is in fact avoided by Peacocke's view in *A Study of Concepts* by stipulation of a regulative role for the theory of reference determination; any set of possession conditions that does not settle on a coherent semantic value is stipulatively ruled out — see (Peacocke 1992, pp.20-21); later Campbell has a similar qualification, see (Campbell 2012, pp.11-12)

ing of the concept is fully given? There is also, however, a somewhat different question that can be asked, about how psychologically realistic this view of concepts can be. What, the worry is, could possibly compel us to use a given concept in the systematic and coordinated ways that we do, if not because of facts about what the concept refers to and how it refers to it? As Peacocke himself later notes, talk of the role simply being 'primitively compelling' seems to leave something to be desired.[22] So we can organise these general worries with this view around two explanatory poles: the causal explanation of our adoption of certain patterns of use, and the normative explanation of the correctness of those patterns.

Perhaps there will be ways of countering these worries. The aim of this section, however, is not to assess the viability of a conceptual-role-first theory of concepts in general, but its viability with respect to the first person concept. And here there seems to arise a problem of a rather more particular kind.

The problem is that the normal patterns of use that we make of the first person concept, made up of the canonical inputs and outputs described in §1.3.2, underdetermines its mode of reference determination. It does not by itself suffice, that is to say, to determine the rule of subject-reflexive reference as the first person concept's mode of reference determination. To see this, consider again what I called the demonstrative model of first person reference in §1.3.1. On that model, uses of the first person concept refer not *de jure* to their producers, but *de jure* to the person to whom the thinker bears the right kinds of perceptual-epistemic relations. The view was given a short hearing. Regardless of it's independent viability as an account of first person reference, however, the important point here is that it would seem to be a fairly simple matter to show its compatibility with the bases and consequences of uses of the first person concept making up its conceptual role. A defender of that view could simply stipulate that the perceptual-epistemic relations that we should be interested in — the ones that determine the reference of our first person thoughts — are the ones that line up in the right way with those epistemic bases and those behavioural implications. There seems to be no difficulty in so reconciling this alternative model of first person reference determination with the conceptual

[22](Peacocke 1998, pp.102-3)

role we have identified for the first person concept.[23]

The conceptual role by itself, then, cannot serve to determine the correct mode of reference determination for the first person concept, since that role is consistent with multiple models of first person reference determination. It won't help to reopen the question whether the demonstrative model should be treated as a feasible competitor to the self-reference rule model. The problem here is that the first person concept's conceptual role underdetermines its mode of reference determination. That is a problem for the role-first order of determinative explanation no matter what we take the correct model of reference determination for the first person concept to be.

1.4.2 Rule determines role

On the second way round of putting things the mode of reference determination for a given singular concept determines its conceptual role, or the normal patterns of use that we make of it. There is, I think, strong plausibility to this picture, coming from the thought that the use that we make of something will be governed by facts about what it does, and how it does it. What else should we *expect* to explain facts about the way we use a given concept, other than facts about what the concept refers to, and how it refers to it.

A different source of motivation for this direction of explanation is offered by a yet later Peacocke in *Truly Understood* (2008):

> Judgement aims at truth (at least), and the norms for a concept seem always to be norms that promote this aim in the case of judgements with certain contents containing the concept. Now the truth or falsity of a judgement turns on the properties and relations of the references of the concepts that feature in the content of the judgement. It is, then, only to be expected that the norms that promote true judgements get a grip at the level of reference.(Peacocke 2008, p.72)

The idea, I take it, is that the sorts of movements in thought mentioned in the conceptual role are those that aim at the truth, or truth preservation; the disposi-

[23]This compatibility is hardly surprising. Even if the demonstrative model is not particularly popular, it would be flat out untenable if it was incompatible with the patterns of use that we in fact make of the concept.

tion to form first person judgments on the basis of epistemic grounds of certain kinds, for example, or the aptness of certain inferential patterns involving first personal premises. Whether or not they are well-designed to succeed in their truth-directed aims, however, depends on what uses of the concept refer to. It is a natural extension of this way of thinking to suppose that the felicity of appealing to *these* transitions and not others in specifying the conceptual role for a given concept will have to await determination of what its uses refer to.[24] [25]

We have already seen this to be a compelling explanatory ordering in the case of logical constants (§1.2), but perhaps a second example will help to further highlight the appeal of this view. Take the perceptual demonstrative concept *that*. Uses of *that* are determined to refer, at least according to the plausible view given above by Dickie, to the object to which the subject stands in an appropriate perceptual-attentional relation. On such a view its conceptual role — as we have already seen in §1.3.2 — is characterised by its basic deployments made on the basis of perceptual attentional relation to an object on the one hand, and the kinds of canonical actions motivated by the holding of such a relation on the other. Seeing, hearing, smelling, tasting or feeling an object in a relevantly attentive way all suffice to put one in a position to have a *that*-thought about it without drawing on any additional information about the situation. On the basis of having such a thought one might, in turn, move one's arm to reach out to it, or one's foot to step on it, or one's whole body to come closer to it as is called for by one's background motivational states.

What explains what in the case of the demonstrative concept? On the face of it, it looks as if it would be mysterious indeed if it turned out that we had settled on the canonical patterns of use associated with demonstrative thought

[24] This argument could be run through in terms of knowledge-preservation rather than truth-preservation without any significant changes for our purposes.

[25] A second initial motivation given by Peacocke in *Truly Understood* is that he take this direction of explanation (or, more precisely, the explanatory claim that the fundamental reference rule for a concept will feature essentially in an explanation of norms distinctive of it) to fall out from two claims many will already accept: that 'the essence of a concept is given by the fundamental condition for something to be its reference', and second that 'there are reasons or norms distinctive of a given concept, where these reasons or norms depend upon the nature of the concept.' (Peacocke 2008) p.54. I am attracted to both of these claims, but will attempt to set out the issues in this chapter without reliance on any particular view about what fundamentally individuates concepts.

ahead of a determination of what it is that uses of the concept refer to. Such a suggestion revives some of the worries of §1.4.1. What could explain our resolution on *these* patterns of use, if not because they characterise a way of thinking of objects that depends on a perceptual relation holding between thinker and referent? What, in the absence of that fact about the conditions of reference determination, would have guaranteed that these patterns of use would give rise to a pattern of semantic value ascription coinciding with coherent and unified referents?

The reverse order of explanation looks far more promising. The reason we use the concept *that* in the way that we do, on this second way of putting things, is because of the way it works to refer. In thinking of an object under a *that* concept, I think of it in virtue of the perceptual relation in which I stand to it — it is a way of thinking about the object that is available to me whenever that relation holds. It is no great surprise, then, that the receipt of perceptual information of the right kind is apt for an immediate application of the *that* concept; once I am in such a situation I need draw on no additional information about the situation before I can think of the object as *that*.

The view that the mode of reference determination is explanatorily prior to its conceptual role is one that is clearly articulated and defended as part of Peacocke's more recent theory of concepts, as given in *Truly Understood* (2008), and employed in *The Mirror of the World* (2014). 'What', he asks in *Truly Understood*, 'is the relation between the rule that gives the reference of a concept and the reasons or norms for making judgments that are distinctive of that concept?' His response — to which he dedicates several chapters of the book — is that 'the rule that gives the reference contributes essentially to the explanation of the norms or reasons specific to the concepts.' (Peacocke 2008), p.53. For Peacocke, the canonical patterns of use that we make of a given concept are explained, at least in part, by the rule specifying the conditions of reference for uses of the concept.

However attractive in general this order of explanation, however, there is a special problem with it in the case of *I*. The problem is that the self-reference rule — the rule that uses of *I* refer to their thinkers, or producers — seems much too modest to account for the rich and complex conceptual role associated with the

first person concept. We have many different ways of grounding first person judgments, some of them direct, some indirect. In all of them, no matter how they are grounded, the resulting thought is one in which a subject thinks of herself in a way that is governed by the self-reference rule. There seems to be nothing in that rule, the rule that uses of *I* refer to their users or producers, that could serve to discriminate between those different ways of grounding first person thought — nothing in the rule itself could be used to separate the direct uses of the concept from the indirect ones. Unlike in the case of the perceptual demonstrative concept, there seems to be nothing special about the mode of reference determination that could straightforwardly guide the subject's use of the concept to be in accordance with its canonical conceptual role.

This sort of difficulty for a reference-first view of the first person concept has been most thoroughly worked out by Campbell. 'The trouble', he writes, 'is that it is very difficult to see how the token-reflexive rule can play the kind of role that the determiner of reference plays in [other cases]. It does not, on the face of it, explain, causally or normatively, the pattern of use that is made of the term.' (Campbell 2012, p. 6). Campbell offers two arguments associated with this claim.

The first comes in the form of a thought experiment. Campbell imagines a sceptic who uses his visual system in the normal way to gain information about the world around him. He uses that information to guide his action, moreover, in just the same way as our use of visual information guides ours. This sceptic fully accepts that the first person concept (or, for Campbell, pronoun, since his interests are at the level of language) is governed by what we have been calling the self-reference rule. What Campbell's sceptic is sceptical about is the possibility of using a concept (or pronoun) whose reference is so determined to directly respond to knowledge gained in the visual way. It is important to note that as we are asked to imagine it, the case is not one in which the sceptic distrusts his senses. He's perfectly happy to accept that the world is as his visual system presents it. He merely doubts that vision is a way of finding out anything about *his own position* in the world. Campbell likens the sceptic's position to that of someone watching a film, or playing a video game: he uses vision as a way of building up a picture of the spatial environment he is viewing, and

even (as in a video game) uses that spatial information to guide his action. It's just that he refuses to treat those informational deliverances as telling him anything about *himself* — he refuses to use the first person as a way of articulating the knowledge thereby gained. As Campbell explains, '[h]e demands some justification for doing so and can find none. The question is whether we can find it for him.'(Campbell 1994), p.119. Campbell is pessimistic about our prospects for doing so.

Earlier, in §1.4.1, I mentioned two aspects of the conceptual role for the first person concept that would seem to be in need of explanation: a normative explanation of its correctness, and a psychological-causal explanation of our conformity with it as competent users of the concept. There are, I think, two ways of reading Campbell's thought experiment, corresponding to scepticism relating to each of these two explanatory demands. The first way of reading it is as asking whether we can find any *theoretical* justification for our sceptic. As theorists of first person thought, do we have the theoretical resources to spell out for the sceptic the correctness of this way of using the concept given its mode of reference determination? The second reading is psychological. Is there any way of convincing the sceptic that, given the first person concept's mode of reference determination, he ought to be moved to make use of it in accordance with its canonical conceptual role? In both cases, it is difficult to see what resources we have to offer the sceptic — there seems to be nothing contained within the self-reference rule that could provide theoretical justification for the interlocking of visual information with uses of the first person concept, and it is surprisingly difficult to spell out to the sceptic why he ought to be moved to use a concept governed by such a rule in response to his visual impressions.

One might respond to this argument by insisting that the thought experiment is just incoherent. Perhaps it is simply inconceivable, when we really think about it, that the sceptic could use vision in the normal action-guiding way, but fail to be disposed to respond first personally to its deliverances, or, at least, that insofar as the case *is* conceivable, his faculty of vision must be importantly unlike to our own. Or perhaps, it might be argued, the sceptic's failing is one of rationality; in refusing to respond to visual information with first person thought he is revealing nothing more than his own rational defi-

ciencies. But even if there is something right in these objections, they do not quite hit the mark — or rather, they simply serve to underscore the force of the problem. The point is that we do not seem to be in a position to *explain* this incoherence or irrationality, even if such incoherence or irrationality there is.[26]

Campbell's second argument can be understood as an attempt to diagnose the special challenge in providing a causal or normative explanation of the first person concept's conceptual role from its mode of reference determination. The problem, as he sees it, is that unlike the perceptual demonstrative case, there is no epistemic relation built in to the mode of reference determination, only the causal relation of *being the thought's thinker*, or *producer*. Given this — and unlike the perceptual demonstrative concept — the rule of reference cannot be a source of guidance (either causally or normatively) about which forms of self-knowledge are apt for a response with its use, or which the forms of behaviour that are thereby made appropriate. As Campbell writes, 'It does not provide one with a way of finding either the bases for making first-person judgments or the consequences that one draws from those judgments.' (Campbell 1994, p. 110).

Despite the general attractions of the reference-first view, then, it looks as if it cannot fully satisfy in the case of first person thought. This is because it looks as if it is not possible to derive the complex patterns of use that we make of the first person concept from the modest resources provided by the self-reference rule, that all uses of *I* refer to their thinkers.

[26]There is a documented neurological patient, DP, who might give us reason to hesitate in giving out these verdicts of incoherence or irrationality too quickly. DP is a sufferer of *anonymous vision*, a condition seemingly caused by hypermetabolisation in the inferior temporal, parieto-occipital and precentral regions, in which visual information is derived from the patient's external environment in the normal way, but an inferential step is needed to for them to recognise that it is *they, themselves* who is observing the scene. DP's symptoms were reported as follows: 'when looking at or concentrating on a new visual object, he is able to see the object as a single object, but that the way he perceived things had markedly changed in a way which he had never experienced before. It appeared to him that he was able to see everything normally, but that he did not immediately recognize that he was the one who perceives and that he needed a second step to become aware that he himself was the one who perceives the object.' (Roland Zahn and Ebert 2008, p. 398) These symptoms share a striking resemblance with parts of Campbell's description of his own sceptic. DP demonstrates, at least, that it is not *inconceivable* that one's visual experiences could be such that one does not immediately recognise them to be apt for response with a first person thought, and likewise there would surely be no failure of rationality on DP's part if he were to take these experiences at face value.

1.4.3 Rule and role are co-determining

A third suggestion accepts that it is a mistake to say that either the mode of reference determination for the first person concept or its conceptual role is more primitive than the other, but holds instead that they are mutually determinatively explanatory. This is Campbell's early view, set out in *Past, Space and Self* (1994). The picture he offers is one on which 'the norms of conceptual role and the norms of semantic value reciprocally regulate one another and can be justified or criticized by appeal to one another.' (Campbell 1994, p.136). This reciprocal regulation is obtained by appeal to a principle Campbell labels *concord*:

> The bases on which judgments using a singular term are made must yield knowledge of the object assigned as reference. (p.137)

Whether we start with the conceptual role or the mode of reference determination, the idea is, concord will provide us with direction about how we should understand the other. Given that the bases on which judgments using the term (or concept) are made must yield knowledge of the object assigned as reference, starting with a certain conceptual role for the concept will constrain which objects can be counted as referent. It can only be the object of which one gains information through the epistemic channels making up that conceptual role. Likewise, if we begin with an understanding of how uses of the concept have their reference determined, concord dictates that the bases of judgments containing that term, the input side of the conceptual role, must be of a kind guaranteed to provide one with knowledge of the object assigned as referent. Neither patterns of use nor rule of reference has explanatory priority; rather, each constrains the other.

Campbell takes this to be an especially well-designed solution to the question of what the relation is between the mode of reference determination and the conceptual role in the case of the first person, and extrapolates to other cases from there. The problems with this picture, however, fall out directly from the discussion of §§1.4.1-2. Even if it seems right that a constraint like concord could ensure the compatibility of the mode of reference determination

and the conceptual role, the enlarged aim of this chapter was to show how one of these could be *derived*, or *determined* by the other. If the discussion of the last two sections is in good standing — if it is right that both aspects of the first person concept underdetermine the other — then it is not at all obvious what added benefit there will be to holding both of those directions of determinative explanation at the same time. If it is correct that each side underdetermines the other, then this condition of mutual accord cannot explain why it is *this* rule of reference determination, given this conceptual role, or *this* conceptual role, given this rule of reference determination.

1.5 The explanatory task

The aim of this chapter was to draw out a double dissociation between the first person concept's conceptual role and its mode of reference determination. In the first direction, the resources contained in the self-reference rule alone seemed too modest to allow for discrimination between the forms of awareness that enter into the conceptual role for *I* and those that don't. In the other direction it seemed as if the first personal conceptual role is equally compatible with more than one candidate mode of reference determination. Both the conceptual role for *I* and its mode of reference determination appear to underdetermine the other.

This raises an explanatory task for theorists of first person thought. An account is owed of how these two aspects of the concept are related that explains either why it is that the first person concept has *this* conceptual role given its mode of reference determination, or *this* mode of reference determination given its conceptual role. Much of the remainder of this dissertation is an attempted undertaking of this task.

Chapter 2

A positive account of first person thought

The last chapter left us with an explanatory task facing theorists of first person thought: if it is right that the canonical conceptual role of *I* and its mode of reference determination by the self reference rule are mutually underdetermining, then what, the question is, explains what? The challenge is to say either why it is that the first person concept has *this* conceptual role given its mode of reference determination, or why *this* mode of reference determination given its conceptual role. This chapter takes on this challenge in the first of these directions — it aims, that is, to find a way of drawing the conceptual role for *I* from its rule of reference determination.

To anticipate, the proposal will be that even if the patterns of use of a concept cannot be derived directly from its rule of reference, that rule engenders certain epistemic conditions on grasping uses of the concept, and it is these epistemic conditions that explain the canonical patterns of use that we make of the concept. This is a fully general proposal with respect to context-dependent singular thought (see pp. 16-17), though I will sometimes drop the qualification 'context-dependent' in what follows. It will be submitted that its application in the case of the first person gives us a way of meeting the explanatory challenge of the last chapter.

The chapter's structure will be as follows. In the first section I identify a certain kind of epistemic achievement that is present in comprehending episodes of singular thought that I call *grasping one's own use of the concept*. I argue for the independence of the conditions of referential success of one's singular thought

from the conditions on this achievement. The central proposal of the chapter comes in §2.2, where it is argued that in grasping a use of a concept, the subject's epistemic state must be brought into alignment with her thought's referential conditions. I apply this model to the case of the first person concept in §2.3, and show how it can account for the canonical patterns of use that we make of that concept. In the final section I make two clarificatory remarks about how the chapter's proposal bears on adjacent discussions in the area.

2.1 Referential vs. grasping conditions

What it takes for a use of a singular concept to refer to an individual is distinct from what it takes to grasp one's own use of it to so refer. This claim describes a thought-theoretic analogue of a phenomenon that shows up quite clearly at the level of language. Take the following example.

> *Sock Drawer:* Getting dressed one morning you are paralysed by your choice of socks, having no reason to prefer one pair over another. Struck by a helpful mood I close my eyes and point at the sock drawer while issuing the instruction, 'go with that pair.'

On a standard view, this deictic use of the demonstrative term 'that' has its reference determined by the condition that it refers to whatever object is gestured at, or made otherwise salient by the speaker. As a competent language-user, let's say I know this to be the case, and insofar as I do there is a sense in which I understand what I am saying in using the word. There is also a sense, however, in which it seems as if you know considerably more than I do about the content of my own utterance: even if I grasp the referential conditions of my utterance, I have no way of identifying the object in the world (if any) that meets them. I am in no position to know any empirical facts about the referent of my thought, even if my thought is referentially successful, because I have no relevant way of identifying the object in question. What this case shows is that there is a difference between the referential conditions of my utterance (which, we may suppose, have been met in this case) and the conditions on a complete understanding of what has been said, or on grasping one's own use of a singular term (which have not).

This distinction — often put as a difference between two different levels of understanding of one's own utterance — has received much notice at the level of language. François Recanati, for example, writes:

> Consider [...] an utterance of '*a* is *G*', where '*a*' is a directly referential term. One 'understands' the utterance in a weak sense when one understands what this utterance linguistically means, namely, that there is an object *x* (possessing a certain property *F*) such that the utterance is true iff *x* is *G*; but to understand the utterance in a stronger sense, that is, to understand the proposition expressed, one must contextually identify the object *x* such that the utterance is true iff it is *G*. The first grade of understanding corresponds to an understanding of the sentence type, a token of which is being uttered; the second grade corresponds to an understanding of what is said by uttering this token. (Recanati 1993)

Similarly, a rich and detailed treatment of the distinction is given by Richard Heck in 'Do Demonstratives Have Senses' (2002). To understand a proposition of the form

> [i]f *U* is an utterance of 'You are a philosopher' and if *x* is the addressee of *U*, then *U* is true iff *x* is a philosopher,

according to Heck, is to grasp the standing meaning of the sentence 'You are a philosopher'. To pass from this to knowledge of what was said by a particular utterance of the sentence, however, the subject must believe *of a particular object* that it meets the condition specified in the antecedent. In doing so, the subject must be in a position to think of the object in question. As he writes, 'if one is to understand an utterance of 'You are a philosopher', one must think of the object that is in fact the addressee in some particular way'. (Heck 2002, pp. 19–20)

Indeed, distinguishing between semantic and epistemic conditions on word- or concept-use might seem like hardly a radical move. What it is to talk or to think about something in a particular way is not the same thing as what it is to fully understand what has been so said or thought, so surely we should only expect that what it takes to *do* these two things will likewise differ. Another look at the demonstrative case, however, but this time at the level of

concepts rather than words, might seem to suggest otherwise, at least at the level of thought.

Unlike the word 'that', the perceptual demonstrative concept *that* does not have its reference determined by conditions involving gestures or saliency-making features of the context. As we saw in the last chapter, rather, *that* has its reference determined by a perceptual-attentional link holding between thinker and object. This difference between thought and language is hardly surprising — in order to express an underlying *that*-thought, a speaker must make salient to her audience which object she is perceptually attending to, and the most obvious way to do this, where it is not evident from context, is with a gesture.

With this reference condition in sight for the concept *that*, however, it becomes difficult to envisage a set of circumstances in which the thinker might successfully refer to an object in thought with a use of *that*, but fail — as in *Sock Drawer* — to be in a position to identify the object in such a way that she could know empirical facts of the form *a is F* about it under the relevant concept. The conditions under which she could successfully think of an object as *that* seem to ensure that she is also in a position to identify the object in the world she is thinking of. After all, how could one successfully think about *that* object by perceptually attending to it, while simultaneously failing to grasp what it is one is thinking of? If it is right that such a separation is not possible at the level of thought, however, then we might hesitate to extend the distinction between the two levels of understanding — or the corresponding distinction between the conditions of reference and the conditions of grasp — at the level of language to the realm of thought.

Another pair of cases, however, appear to reveal the distinction to be just as live in thought as it is in language, even if there are some concepts for which the two sets of conditions always seem to stand and fall together.[1]

> *Kidnap:* You have been kidnapped. As a protective measure, your kidnappers have locally aneasthetised you all over, placed you in a sensory deprivation tank, injected you with amnesia-inducing drugs, and driven you to a secret location. You think to yourself, *I need to escape from here*.

[1]This covariation in conditions of reference and conditions of grasp in the case of *that* is taken up again and explained on pp. 60-61.

The concept *here* has its reference governed by the rule that its uses always refer to its location of utterance. This use of *here* refers to wherever you have been left by your kidnappers. In knowing this much, there is a sense in which you know what it is you are thinking. There is another sense, however, in which you are rather more in the dark about the content of your own thought. This is the sense in which you do not know anything about the place you are thinking of in your use of the concept, or are in no position to know things of the form *it's F here* about that location. (You are receiving, let us suppose *ex hypothesi*, no information whatsoever from the location you are currently in). As I have been using the phrase, you are not in a position to *grasp your own use of that concept*.

A similar case can be constructed for a use of the concept *today*.

> *Rip:* Upon awakening, Rip Van Winkle has no idea how long he has been asleep. He quickly realises, however, that things are even worse than he had first thought. He finds that he has been anaesthetised all over, submerged in a sensory deprivation tank, and pumped full of drugs that both induce amnesia and distort his perception of passing time. Desperate to get his bearings he thinks to himself, *I wonder if today is a Tuesday....*

His use of the concept *today* refers to the day on which his thought episode occurs. In using the concept competently, moreover, Rip knows as much. There is also, however, a kind of epistemic failure here. He does not know, in the relevant sense, anything about the day he is thinking of; he is in no position to know empirical facts of the form *it is F today* about it. He fails, that is to say, to accomplish the cognitive achievement that I have called *grasping one's own use of a singular concept*.

2.2 The proposal

There is something missing from both *Kidnap* and *Rip*. In both cases the subject thinks a thought about something (a place, a day), but fails in an important sense to fully understand what has been so thought because of a failure to be epistemically related to the thing (place, day) being so thought about. Actually, what is missing is more specific than this. It's not merely that the subject

fails to be epistemically related to her object of thought. It's that she fails to be epistemically related to it *in the right way*.

To see this, notice that not just any epistemic relation will do. In *Kidnap*, for instance, let's imagine that your kidnappers find a way to transmit signals directly into your visual cortex from a year-old video of your current location. Under the right conditions, this might be a way of knowing about the location that is, in fact, the place where you are. Even if that is right, however, this epistemic link to your location does not seem to be of the right kind to secure your comprehension of your use of the concept *here* — it doesn't have the same impact on your *here*-thought as the opening of my eyes would have had on my *that*-utterance in *Sock Drawer*. Likewise, the faint memory breaking through Rip's amnestic fog that the 30th December 2014 is a Tuesday does nothing by itself to help him grasp his use of *today*, even if that is, in fact, the day about which he's thinking. What is missing in both of these cases is a particular kind of epistemic relation: namely, one that aligns with the indexical form of those thoughts.

Suppose now that instead of transmitting images into your visual cortex, one of your kidnappers cuts a small hole in your blindfold, giving you minimal but clear access to visual information from your immediate environment. You think again to yourself, *I really need to escape from here*. Unlike your last thought, your use of the concept *here* really does now seem to be flooded with comprehension. You now are now in epistemic contact with the place you are thinking of in a way that allows you to grasp your own use of the concept *here*; you know about the place you're thinking of, such that you are in a position to know empirical facts about it like *it's bright here*, or *it doesn't look like there are any windows here*. There is a sense in which you are now in a position to understand what it is that you just thought, in a way that you weren't able to before.

Similarly in *Rip*, suppose that the drugs start to clear. At first, we may suppose, Rip begins to feel just a few faint sensations in his numbed limbs, to remember a handful of faded memories, and little by little to regain control of his capacity to track the passing of time. But soon his head is clear enough, and after contemplating the passing of time for what he judges to be about a day, he thinks again to himself the thought *I wonder if today is a Tuesday*. The

content of the thought is the same as the one he had entertained before. But — like in the amended *Kidnap* case — it now seems permeated with a new level of comprehension. He now is has ways of knowing about the day he is thinking of, of a kind that he didn't have access to before; he might go ahead and form the judgment, for instance, that *it's cold in here today*, or *time seems to be passing slowly today*. He is now in a position, that is to say, to grasp his own use of the concept *today*.

What is it that makes *these* epistemic conditions apt for grasping uses of the concepts *here* and *today* (the hole in the blindfold, the clearing of the drugs), and not the ones described three paragraphs ago (the visual projections, the foggy propositional memory)? The proposal of this chapter is that in being related to their immediate environments in the ways described in the newly amended cases, these thinkers are now epistemically related to their objects of thought in ways that line up with their ways of thinking of them; they are, as we might say, epistemically inhabiting their referential perspective on the objects of their thoughts. Or, to put it another way, the thinker's epistemic relation and her referential relation to the object of her thought are both constrained by one and the same overarching constraint.

Take first the amended *Kidnap* case. The mode of reference determination for *here*, we saw, is given by the location-reflexive rule that uses of the concept always refer to their locations of use. The overarching condition that constrains this form of thought about a place, then, is that of *identity of location*; the place thought about must be one and the same as the place at which it is thought about if one is to think of it as *here*. The present proposal is that it is this very same overarching condition that must constrain the thinker's epistemic state too if she is to grasp her own use of the concept. She must, that is to say, be epistemically related to the location in a way that essentially depends on the location from which the information is derived being the same as the location at which the information is received if she is to latch on to the object of her *here*-thought in a way that suffices for full comprehension. Vision is just such an epistemic relation. This explains why cutting a hole in the blindfold allows you to grasp your own use of the concept *here* in a way that you couldn't before.

A similar story can be told for the amended *Rip* case. The mode of reference

determination for *today* is given by the rule of day-reflexive reference, that any use of it refers to the day of its use. As before, then, we can identify an overarching condition on uses of *today*, which is that of identity of day. The day on which the thought is thought must be the very same one as the day being so thought about. Now if Rip is to grasp his use of this concept, my suggestion has been, then he must be in a position to epistemically take up the same sort of relation to his object of thought as is manifested by the thought's referential conditions. He must be epistemically related to the day in question in a way that is restricted by the very same condition as the one constraining the reference of his thought. Given that that condition is one of day-identity, this means that he must be epistemically related to the day in a way that depends on the sameness of the day from which the information is derived, and the day on which it is received. The kind of epistemic capacities Rip regains once the drugs begin to clear — proprioceptive information, or an introspective capacity to track the passing of time — are of just this kind. They are ways of knowing about how things are on a particular day that depend essentially on the identity of the day known about and the day on which the knowing is done. It is because of this alignment between Rip's new epistemic circumstances and the referential conditions on his *today*-thought, or their mutual restriction by the overarching condition of day-identity, that explains why Rip is in a position to grasp his own use of the *today* concept in the amended case, but not before.[2]

The proposal, then, is this. In order to be in a position to grasp one's own use of a concept, a subject must align her epistemic and her referential states. She must be related to the object of her thought through a potentially knowledge-providing relation that follows the same trajectory, or is constrained by the same overarching condition, as the referential relation holding between her use of the concept and its referent.

Less freely put, the central proposal of this chapter is captured by the following two principles.

[2] As is turns out, we will see in §2.3 that *Rip* actually might have been in a position to grasp his use of *today* all along — on a certain view of introspection, his introspective access to the temporal properties of his thoughts would be enough to meet the conditions on concept grasp for a *today*-thought. I put this complication to one side for the moment; I hope the example still serves to help to bring the relevant phenomenon to view.

> *Concept-grasp:* A thinker uses a singular concept *c* comprehendingly, or equivalently *grasps her own use* of *c*, if and only if she is in a position to know empirical facts of the form *a is F* about the object under *c*.

> *Mutual constraint:* A thinker is in a position to know empirical facts about an object of the form *a is F* under *c* only if she is epistemically related to the object in a way (or ways) that are constrained by the same overarching condition constraining the referential conditions of her use of the concept *c*.

Four clarificatory points are immediately in order. The first is a word on the phrase 'in a position to know' that connects the two principles. There are many ways in which a subject might be barred from knowledge of the relevant kinds of empirical facts, even if the perceptual system through which such knowledge would normally be gained is physiologically fully operational. Like Campbell's sceptic from §1.4.2, for instance, the subject might be epistemically aversive and refuse to form beliefs on its basis, or fake barn type scenarios might prevent a subject from gaining knowledge about the relevant object for reasons outside her control. These are not the sorts of obstacles to knowledge that are relevant to the principles *concept-grasp* and *mutual constraint*. Talk of a subject's being 'in a position to know' certain things about an object should be understood as talk of what an epistemically virtuous agent could know about the object in epistemically non-deviant conditions.

Secondly, the proposal is that there is a kind of cognitive achievement definitionally captured by *concept-grasp* that is normally present in episodes of singular thought. This achievement is different from the referential success of the thought, but makes a difference to whether the episode can be counted as a comprehending one (in the relevant sense). The principle does not arbitrate on whether there will be other forms of cognitive achievement in the area worth investigating. It is left open, for instance, that there might be a number of different degrees of comprehending singular thought characterised by different kinds of achievement, only one of which corresponds to *concept-grasp*, or that there might be other kinds of achievement characterising alternative dimensions altogether along which a thought can be said to succeed or fail.[3]

[3]Someone else who has tried to make a case for the existence of different 'levels' of reference,

The third thing to say is that *mutual constraint* is intended as a threshold condition on grasping one's own uses of a concept. As long as the thinker is epistemically related to her object of thought in any way at all through a relation constrained by the same overarching constraint that applies also the referential conditions of her thought, then she is thereby in a position to grasp her use of the concept. There may be many such epistemic relations for a given kind of overarching constraint, or there may be just one — and even where there are many, the thinker need not exploit them all. Variation in the epistemic relations meeting the condition specified in *mutual constraint* and exploited by the thinker in any given episode of singular thought will affect her overall psychological state in grasping her use of the concept, but as long as she exploits just one such epistemic relation, such variations can make no difference to whether she can be said to be grasping her use of the concept at all. Concept-grasp of the kind captured by the above principles, the suggestion is, does not come in degrees.

A fourth quick point to note is that it should be clear that the epistemic relations mentioned in *mutual constraint* are relations that a thinker must *be in* in order to grasp a use of her own concept; she need in no way represent them as part of the thought, nor have the conceptual capacities to do so.

Earlier we saw that the conditions on reference and the conditions on concept-grasp did not seem to come apart in the case of the perceptual demonstrative concept *that*.[4] We are now in a position to see why this is. Even before appealing to an overarching constraint on the thinker's referential and epistemic relations to the object, in the case of *that* there is no chance that the referential conditions on one's thought can be met without the epistemic conditions on concept-grasp being met too. That is because the relevant epistemic relation is, in the case of *that*, *built in* to the conditions of reference — the condition, that is, of standing in a perceptual-attentional relation to the object. There is no possibility of meeting this condition without also meeting the epistemic conditions on concept-grasp, because to meet the first condition is already to meet

depending on the presence or absence of epistemic rapport of a certain kind with the object is Imogen Dickie; see (Dickie 2011a).

[4] p.54

the second. It is only in the special case of concepts like the pure indexicals *today* and *here*, where no epistemic relation is mentioned in the conditions of reference, for which the referential conditions and the conditions on comprehending concept-use are conceivably separable. That is because to meet the conditions of reference in these cases is not yet to automatically meet the conditions on concept-grasp.[5]

It is, at this point, perhaps worth remembering that it is because of this very same feature of these pure indexical concepts that it is with them that the challenge of chapter 1 really found its bite — for it is only when there is no epistemic relation specified in the reference conditions that it begins to look mysterious why it is that a given concept has a particular pattern of use, given its rule of reference determination. It is no coincidence that the conceivability of the conditions on reference and on concept-grasp coming apart makes these concepts particularly well-suited both to bring out the problem from the last chapter, and the solution given here. An overarching constraint on the referential and concept-grasp conditions for a concept like *that*, where meeting the first set of conditions suffices for meeting the second, will be trivially met. It is only in the case of pure indexical concepts like *here*, where the two sets of conditions need not always be met together, that the presence of this overarching constraint is brought into the foreground.

Another such case, of course, is *I*.

2.3 First person thought

What should we say about the first person concept? We already have in place its rule of reference determination — the rule, that is, that uses of the concept always refer to their thinkers. The overarching condition on first person reference, then, is again that of identity, but this time it is identity of *subject*: the subject who uses the concept must be the same as the one to whom it refers if she is to refer to that person under the first person concept *I*.

[5]We will see in the next section that there is a view of introspection that might undermine this conceivable separability in the case of *today*. But this will be for a reason other than the epistemic conditions getting built in to the referential conditions.

From here we are ready to employ the proposal of the last section. According to *mutual constraint*, grasping one's own use of a singular concept requires that one be suitably epistemically related to one's object of thought. One must, under that principle, be epistemically related to it in a way that is constrained by the same overarching condition that constrains the referential trajectory of one's thought. In the case of *I*-thought this is to say that in comprehending uses of the first person concept one must be appropriately epistemically related to oneself.

Not just any epistemic relation to oneself, of course, will do. Seeing live CCTV footage of oneself, for instance, would not help in grasping one's own use of a first person concept; although it is *a* way of bearing a potentially knowledge-providing relation to the right object in the world, it is not a way of the right kind. According to the proposal of the last section, rather, one must be epistemically self-related in a way that aligns with the referential trajectory of the *I*-thought, or in a way that is constrained by the same overarching condition. That condition is subject-identity. Applied to the case of the first person concept, then, the proposal of the last section comes to this: that in order to grasp one's own use of the first person concept, *one must be related to oneself through an essentially subject-reflexive epistemic relation.* One must be related to oneself, that is to say, through an epistemic relation (or relations) whose functioning essentially depends on the identity of the knowing subject and the subject known.

This account of what it takes to grasp one's own use the first person concept arms us with the resources to return to the explanatory challenge left unanswered at the end of the last chapter. Is there, the question was, a way of saying why it is that the first person concept has *this* conceptual role, given its mode of reference determination by the self reference rule?

We are now in a position to give the following answer. Even if the canonical patterns of use that we make of the first person concept cannot be directly explained by that concept's mode of reference determination by the self reference rule, that mode of reference determination gives rise to certain epistemic conditions on comprehending uses of the concept. In order to grasp one's own use of the first person concept, one must be in a position to exploit forms of self-knowledge of a very particular kind. One must, in conformity with the

principle of *mutual constraint*, be in a epistemic contact with oneself through forms of self-knowledge that are essentially subject reflexive — not because those channels of self-knowledge serve to fix the reference of our first person thoughts, but because they enable comprehending thoughts of that kind. It is these epistemic constraints on grasping one's own use of a first person concept, then, that explains why the first person concept has the conceptual role that it does. And given that the rule of reference for *I* explains why these epistemic constraints are the ones that they are, this is to say that the mode of reference determination for *I* ultimately allows us to say why it is that the first person concept has the conceptual role that it does.

Another way to put the proposed response is to say that even if the reference rule for *I* is, by itself, indiscriminate between different forms of self knowledge — and so, by itself, cannot say why it is that some such forms get into the conceptual role for *I* and not others — the epistemic constraints that it generates on comprehending uses of the concept are not so neutral. On the present account, it is these epistemic conditions on comprehending first person thought that explain the canonical moves that we make with the concept *I*, rather than the rule of reference determination that gives rise to them itself. The first person concept can be used to directly articulate certain forms of self-knowledge and not others, because it is these forms of self-knowledge, and not others, that ground the subject's comprehending use of the concept in the first place.[6]

What might these forms of self-knowledge be? We might take as an example the normal way of knowing, when one does, of one's own blushing from the inside — a form of self-knowledge that depends essentially on the identity of the knower and the known.[7] It is a way, that is to say, of knowing about my own blushing and no one else's, and what's more, such subject-reflexivity is an

[6] This proposal has clear affiliations with the account of first person thought offered by Rödl: 'we inquire after the sense of "I", the logical perspective that this form of reference affords on an object. I refer to myself first personally, not through an individuating concept, but through a relationship with the object by which I know how things stand with it. Since first person reference is reference as to oneself, the relation is identity. First person knowledge is knowledge one has not by perceiving but by being its object. In order to explain the sense of "I", we must describe this form of knowledge.'(Rödl 2007, p. 10)

[7] As an illustrative example, I put aside for the moment any putative counterexamples involving causal deviant chains, but see chapter 5 for a serious defence of the point.

essential, or *de jure* feature of this epistemic relation. If *mutual constraint* is in good standing, then, this will be one the forms of epistemic self-acquaintance through which I will be in a position to grasp my own use of a first person concept. The proposal of this chapter, then, is that by knowing about my own blushing from the inside, I will be in a position to grasp my own use of *I* in thought; I will be able to know — without drawing on any identificatory beliefs — empirical facts of the form *a is F* (I am blushing) under the relevant concept (*I*), about my object of thought (me). If this is right, then it is no wonder that it is part of the normal pattern of I make of the concept *I* that I am apt to use it in immediate response to knowledge gained in this way that I am blushing. That this is a form of self-knowledge that can ground my comprehending use of the concept explains why this form of self-knowledge enters into the conceptual role for *I*.[8]

Beyond this sample example, I owe, of course, a more detailed account of what these forms of essentially subject-reflexive self-knowledge are. In chapters 4 and 5 I will argue that they include, at least, introspection, internal bodily awareness, multimodal bodily awareness, and episodic memory. In arguing for this range I am attracted to the idea that our first person way of thinking of ourselves is grounded in physical self-knowledge as well as mental. In this I am a materialist about self-conscious thought, insofar as that doctrine is characterised by the claim that we are aware of ourselves, as ourselves, as physical as well as mental entities. That there is nonetheless space on the proposed account for a certain kind of prioritisation of mental self-knowledge is, however, worth bringing out.

As Anscombe made vivid, it is perfectly conceivable that we could suspend a thinker's access to information about her current physical states in the con-

[8]In chapter 1 I distinguished the input and the output sides of the conceptual role, and said that I would concentrate on the former in this dissertation for reasons of scope. But *concept-grasp* and *mutual constraint* together also give us a way of explaining why *some* forms of action become available upon the deployment of a first person concept, and others don't. Together these principles ensure that whenever a thinker comprehendingly thinks a first person thought, she will be in a position to know about herself in a way, or ways, that essentially depend on the identity of herself *qua* knower, and herself *qua* object known. The forms of action that will become newly available, then, are those that are appropriate *given that identity* — they will be, as Perry calls them, 'normally self-dependent/directed/effecting action[s]', rather than those that would be appropriate where the knower and the known are distinct subjects. (Perry 1990)

tinued presence of a capacity for first person thought — and, we might add, a continued capacity for *comprehending* first person thought, in which the thinker grasps her own use of the first person concept. We have already seen the case in chapter 1, but here it is again:

> 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!" (Anscombe 1997, p. 153)

As this case illustrates, the ability to think about oneself first personally, and in doing to so to grasp what one is thinking, seems not to depend on a thinker's ongoing bodily self-awareness.[9]

The interest of Anscombe's sensorily deprived subject for our purposes is to make salient an arguable point of contrast between our bodily and our mental forms of self-awareness. While there seems to be no conceptual obstacle to the possibility of meaningful conscious first person thought in the absence of bodily self-awareness, it is not clear that the same can be said of our mental self-awareness. Unlike the physical case, it is tempting to wonder what it would really *mean* to have a conscious first person thought — or, indeed, a conscious thought of any kind — to which the subject could in principle have no introspective access. Unlike the case of bodily perception, what it is that one is aware of in undergoing a conscious thought *just is* the conscious thought itself. There seems, then, to be no possibility of undergoing that thought in the suspension of one's introspective faculties; as long as a thought is conscious, it must be one to which the subject has at least the potential to access. *A fortoriori*, then, all conscious first person thought must be at least potentially introspectively accessible thought.[10]

[9]There is, as we have already seen, no immediate tension between this intuition and the present account of first person thought, on which it is submitted only that the thinker must be related to herself in *some or other* essentially reflexive form of self-knowledge in order to grasp her use of the first person concept. Not all such forms of self-knowledge need be available for all comprehending uses of that concept.

[10]Three notes of caution on this point: (i) It does not commit us to anything as strong as the

If this is right, then there is an important respect in which it brings first person thought alongside perceptual demonstrative thought: it now looks as if there is no possibility of having a successfully referring use of a first person concept without the thinker also being in a position to grasp it. The suggestion of the last paragraph was that all conscious first person thoughts will be had in the context of introspective self-awareness (because all conscious thoughts are). But if introspective self-awareness is one of the forms of self-knowledge underpinning grasp of our uses of the first person concept, then this means that all conscious first person thoughts will be had in a context within which the use of the first person concept can be grasped by the thinker. Just like in the case of the concept *that*, the conditions on reference and the conditions on grasping one's own use of the concept seem to stand and fall together.

There is, however, a crucial difference between the two cases. In the case of *that*, we saw that it is because the epistemic condition on grasping the use of the concept is written into the referential conditions themselves that the two could not come apart; one could not meet those referential conditions without already having met the relevant epistemic conditions on concept-grasp. This is not the same in the case of *I*, for which there are no epistemic conditions entering into the conditions for reference. The reason that the two sets of conditions are always met together in the case of the first person concept, rather, is because at least one of the forms of self-knowledge that would allow a thinker to meet the conditions on concept-grasp (*viz*. introspection) also stands as a background enabling condition on consciously using a concept at all, first personal or otherwise. This point of resemblance between first personal and perceptual demonstrative thought, then, should not mislead us into giving them parallel treatments.

It is tempting to extend this point in the case of *I* to cases of temporal indexicals too, such as *now* and *today*. If it is right that conscious thoughts are

claim that thoughts must always *in fact* be the objects of introspective reflection by the thinker, only that they must be *potential* such objects. (ii) These comments should perhaps be restricted to the kinds of creatures that we are. Perhaps there are, or could be, more primitive creatures who lack an introspective faculty as we recognise it, but who can still entertain thoughts with first personal mental content. (iii) It also leaves open the possible view that there may be unconscious first person thought for the kinds of creatures we are too, in which the referential but not the grasping conditions are met, e.g. in dreams.

always the potential objects of introspection, then it seems that *now*- and *today*-thoughts will also always be had in a context in which the conditions on grasping the use of those concepts are guaranteed to be met. That's because to grasp those concepts, the thinker must be related to her objects of thought in a way that is constrained by the very same overarching constraint that governs the reference of her thoughts. In the case of *now* and *today*, those will be the constraints of moment-reflexivity and day-reflexivity respectively — and introspective awareness of occurrent thinking provides an epistemic channel meeting both of those conditions. Introspection, then, serves as an epistemic relation that enables concept-grasp for both *now*- and *today*-thoughts (and, presumably, other temporal indexicals too). If this is right then it might force something of a rescindment on the earlier case of *Rip*. That case was introduced to illustrate the possibility of a separation in thought between the conditions of reference and the conditions of grasp. If it is right that all conscious thought is potentially introspectible thought, then it might turn out that it does not give us such a case after all.

What about *Kidnap* — does the suggested pervasiveness of introspective access to conscious thought also show that we could not have conscious *here*-thought in the absence of meeting the conditions on comprehension either? I do not think that it does. Introspection is a temporally but not a spatially structured faculty; it gives us access to temporal but not to spatial properties of our mental activities. This means that to introspect one's own conscious *here*-thought does not put a thinker in position to know empirical facts of the form *there's thinking happening here* in the same way that it allows her to think *there's thinking happening now*. Of course, this is not to deny that whenever there is introspected thought happening now, it is also happening here. It is only to say that the *hereness* of the thought is not delivered up to the subject through introspection in the same way as its *nowness* is. Compare a scenario in which you touch but do not see an irregularly shaped middle-sized object under the table. Your sense of touch gives you a way of knowing about some of the object's extensional properties; you can feel that it is bounded and extended in space, and with a little exploratory movement of your fingers you can also discover something about its shape and size. Given these extensional properties,

it is safe to assume that the object also has properties of other kinds. Among others, it will have weight-properties, surface colour-properties, composition-properties and subtended angle-properties. Given your sensitivity to the dependence relations between these various property-kinds, to feel an object's extensional properties might put you in a position to infer that it also has properties of these other kinds — indeed, given the pragmatic importance of such dependencies it will likely put you in a position to make such inferences with maximal psychological fluency. But it does not directly put you in a position to know empirical facts *about* those other properties of the object. Likewise, introspective awareness of thinking might put you in a position to infer that *there is thinking happening here*, but it does not position you to know empirical facts about the spatial properties of those thoughts. Unlike *I* and *now* thoughts, introspection does not provide the right kind of epistemic channel for grasping one's own uses of *here*.

I have not fully argued here that thoughts can only be entertained in the context of potential introspective self-awareness; I have only observed that if that *is* right, then mental self-knowledge would play a special role on the present account of securing covariation in the satisfaction of the referential and the concept-grasp conditions on first person thought (and, perhaps, temporal indexical thought too). There are at least two explanatory payoffs for someone prepared to argue the point.[11] The first is that it explains why it is that first person thought seems to be ineliminably self-conscious. It explains, that is, the intuition that I couldn't have a successfully referring first person thought without realising that it's *me* I'm thinking of in having it. This is explained, on the present account, by the fact that it is not possible to *use* a first person concept in the first place without already having met a sufficient condition for grasping that use.

The second explanatory profit is that it places introspective self-knowledge in a central position within a materialist account of the self-consciousness underlying first person thought. While holding onto the naturalistic intuitions

[11] See (Schear 2009) for someone who does argue this point, and see work by Uriah Kriegel and Dan Zahavi for two writers who argue for the stronger claim coming out of the phenomenological tradition that conscious thought is always itself an object of awareness for the thinking subject, e.g. (Kriegel 2009b), (Kriegel 2009a), (Zahavi 2005).

underpinning materialist accounts of self-consciousness as found in such writers as Evans, Cassam, Ayers, Rödl, McDowell, Brewer, and P.F. Strawson[12], it nonetheless explains what is right about the seemingly unshakeable idea that there is a privileged connection between first person thought and introspective self-consciousness.

2.4 Two clarificatory remarks

Russell's principle

The notion of being epistemically related to one's object of thought drawn on in this chapter should not be mistaken for anything as strong as Gareth Evans' *Russell's principle*, formulated by Evans as

> The principle that a subject cannot make a judgment about something unless he knows which object his judgment is about (Evans 1982, p. 89).

Evans substantiates this principle with a more detailed account of what it takes to know which object one's judgment is about. To possess this knowledge is — according to Evans — to have *disciminating knowledge* of it; the subject must 'have the capacity to disinguish the object of his judgment from all other things.' (Evans 1982, ibid) There are at least three ways she can do this: by having a direct perception of it; by a capacity to recognise it; or by possession of discriminating information about it.

The requirements on thinkers of singular thoughts laid on by Russell's principle are often taken to be unappealingly heavyweight. A case given by Evans himself suffices to make the point. The case is well-known: Evans asks us to imagine a subject who one day observes a steel ball rotating about a point. On a later day, the subject observes a second steel ball rotating about the same point, a ball that is perfectly indistinguishable in every way from the first. The subject, we are asked to suppose *ex hypothesi*, has no way to discriminate between the two balls. The subject then suffers from local amnesia that eradicates in full the subject's memory of the first ball.

[12]Evans 1982, Chap 7; Cassam 1997, pp. 2–9; Ayers 1991, p. 285; Rödl 2007, 14–15, Chap 4; Strawson 1959, pp. 162–9; Brewer 1995, pp. 297–306; McDowell 1994

Years later, Evans imagines the subject reminiscing about 'that shiny ball'. The subject fails to satisfy Russell's principle; he lacks the capacity to discriminate the ball about which he is essaying a thought from all other things, since he has no way to discriminate it from the first ball. As Evans points out, however, 'it would certainly be quite natural, in view of the facts, to say that he was *thinking of* the second ball' (Evans 1982, p. 90). As Evans freely allows, the verdict returned by Russell's principle is considerably less permissive than intuitive judgment about whether there is successful singular thought here. Possession of discriminating knowledge of an object is a forbiddingly high bar to set on the success of object-directed thought.

The bar on being epistemically related to an object that has been argued in this chapter to fund one's grasp of a use of a singular concept is much lower than this. The subject need not have the capacity to discriminate her object of thought from all other things. Her epistemic rapport with the object need only put her in a position to know empirical facts of the form *a is F* about it under the relevant concept, not to discriminate it from all other things. Under this weaker constraint, there is no reason to think that Evans's subject could not comprehendingly think a thought about the second steel ball under a memory demonstrative concept, *that ball*. According to the proposal of this chapter, he must only be in a position to know an empirical fact about it under that concept, and for that his epistemic relation to the ball must align with the referential conditions of his thought. A natural-sounding way of filling this suggestion out is to say that meeting both sets of conditions (referential and epistemic) the subject must meet a combinatorial overarching condition: the subject must have previously borne the right kind of perceptual attentional relation to the ball, and for the experiential impression thereby created to have been preserved in the right way in memory. There is no appeal here to anything as strong as Russell's principle, and no reason to think that the subject could not meet the proposed weaker condition.

Acquaintance vs. anti-acquaintance theories

The second note of clarification concerns the ongoing acquaintance/anti-acquaintance debate about singular language and thought. How one understands the intersection between that debate and the current proposal crucially depends on how the interests of that debate are formulated. Participants on both sides commonly write as if what is at issue is the question of whether there is an acquaintance restriction on the referential relation involved in singular thought and language. Hawthorne and Manley, for instance, campaign for 'the thesis that reference and singular thought are not subject to any interesting acquaintance constraint' (Hawthorne and Manley 2012), while Recanati characterises the question at the centre of the debate as that of whether an acquaintance condition holds for our 'mental relation to individual objects' (Recanati 2012a, p. 3). So long as the target of the acquaintance/anti-acquaintance debate is understood as the question of whether there is an important category of thought or linguistic expressions for which *referential* success depends on the thinker's acquaintance with her object of reference, then the proposal of this chapter is orthogonal to that debate. The current proposal posits an epistemic constraint not on reference, but on a different cognitive acheivement associated with singular thought — that of grasping one's own use of a concept.

Having said this, however, there is also a way in which the proposal of this chapter might be of interest to anti-acquaintance theorists like Hawthorne and Manley, with the debate so characterised. Argument in this area typically proceeds by the urging of intuitions about whether or not thoughts and utterances produced in the seeming absence of the relevant acquaintance relations can really be said to refer. The proposal of this chapter has been that referential success is not the only important cognitive achievement associated with singular thought. One might succeed in referring to something in thought, but still fail to grasp one's own use of the singular concept. If this is right, then it offers anti-acquaintance theorists a new resource with which to ply our intuitions. In cases in which it seems as if the intuitive verdict is with the acquaintance theorists — where is seems, that is, as if there really couldn't be reference without acquaintance — the anti-acquaintance theorist can encourage us to look again to see

whether the failing might not be one not of reference, but of concept-grasp.

Another way of formulating the acquaintance/anti-acquaintance debate, however, pegs the current proposal as much less friendly to the anti-acquaintance side. The question that concerns the anti-acquaintance theorist might not be the narrower one about reference, but the broader one that asks, is there *any* interesting epistemic constraint associated with singular thought, or language at all? To this broader question my proposal offers a positive answer — yes, there are epistemic constraints on grasping one's own use of singular concepts.

There is some evidence that this broader reading is what Hawthorne and Manley have in mind when they fleetingly consider a suggestion similar to my own. They take the view that there is an acquaintance constraint on *fully understanding* a singular content, rather than on singular content simpliciter, as a potential rival to their position. They quickly put the suggestion aside on the grounds that it would create a troubling possibility of linguistic reference without underlying comprehension, and so without the possibility of conferring comprehension to one's hearers: 'this view complicates a natural picture of communication on which one expects one's interlocutors to understand the propositions one semantically expresses.' (Hawthorne and Manley 2012, p. 24) This might be a fair objection to nearby views, but it would be badly aimed at the current proposal. This, first, is because on the current proposal concept-grasp is a cognitive achievement that is nearly always met, so even if there is the conceptual possibility on this view of utterance without comprehension, it is extremely rare, and second, because cases like *sock drawer* were supposed to show comprehension-failure at the level of language to be something to be made room for, not avoided.

2.5 Conclusion

The question left hanging at the end of the last chapter was, is there a way of saying either why the first person concept has *this* conceptual role, given its mode of reference determination, or why *this* mode of reference determination, given its conceptual role? This chapter has tendered a response to this challenge in

the first of these directions: drawing on the principles *concept-grasp* and *mutual constraint* from §2.2, it offered a way of deriving the conceptual role for *I* from the epistemic constraints on grasping one's own use of the first person concept, which are in turn derived from the rule of reference determination for *I* by the self reference rule.

At the heart of this account of first person thought is the idea that reference is not the only measure of success for singular thought. It takes more to think a first person thought — or, at least, to think it with full comprehension — than just meeting the conditions specified by the rule of reference determination for *I*.

Chapter 3

Immunity to error through misidentification

Some first person judgments are immune to error through misidentification relative to the use of the first person concept. This, roughly, is to say that the judging subject couldn't have gone wrong solely in having misidentified the person to whom she knows the judgment's predicate to apply. Other first person judgments are not. There is a prima facie explanatory challenge generated by this asymmetry that is taken up in this chapter.

The challenge takes on a special force for theories of first person thought under which the reference of uses of the first person concept is determined by the simple rule that uses of *I* refer to their thinkers. That is because a tempting approach to questions of immunity to error through misidentification — and one that, as we will see, proves to be effective in providing explanations of immunity to error through misidentification of other judgment-kinds — is to look to the way in which the reference of the concept to which the immunity is relativised is fixed. As we will see in §3.2, a matching of the relevant kind between the grounds or the content of predication and those of reference-fixing suffices to ensure that the subject could not have made an error in applying that predicate to that referent, so grounded and so determined. But this strategy is unavailable to self-reference rule accounts of the first person concept. That is because we will see in 3.3 that if there is to be a matching between the grounds or the content of predication in first person judgments with immunity to error through misidentification and the first personal mode of reference determination, then the latter will have to be understood either on a demonstrative model,

or as descriptively determined. To explain it by a match between the mode of reference determination and the *grounds* of the predication would have us treating the first person as determined to refer demonstratively to the source of proprioceptive, introspective, etc. information; to explain it by a match between the mode of reference determination and the *content* of the predication would leave us with an understanding of *I* as given by a descriptive condition, like *the person blushing*. Either way, we would be obliged to leave behind the idea that it is determined by the rule that uses of *I* refer to their thinkers.[1]

The result is a distinctive form of the challenge. Given that the first person concept is governed by the rule that uses of the concept refer to their thinkers, how are we to explain the asymmetry between those first person judgments that are, and those that aren't, immune to error through misidentification relative to a use of the first person concept?

The central proposal of this chapter is an expansion of the kinds of explanation that we should take to be available to account for cases of immunity to error through misidentification. In particular, I argue that although there are a range of central cases in which the explanation draws on a matching between the predicative component of the judgment and the concept's mode of reference-fixing, other cases are susceptible to an explanation coming from a matching of a different kind — between the grounds of predication and the grounds of *concept-grasp* on the part of the subject. I go on to show that cases of immunity to error through misidentification relative to uses of the first person concept should be understood as cases of this latter kind.

In §3.1 I set out the epistemological phenomenon of immunity to error through misidentification. I offer three ways to explain its presence in §3.2, each conforming to a more general explanatory template for explanations of immunity to error through misidentification. I take up the asymmetry challenge in §3.3. The right explanation in the case of the first person, I argue, comes not from the conditions on reference-fixing, but from the conditions on concept-grasp. In the fourth section I consider the implications of this form of

[1]The special force of this challenge for self-reference rule accounts of first person thought is also recognised by both Peacocke and Campbell; see (Peacocke 2008, pp.96-7) and (Campbell 2012, p.13).

explanation for the significance of the opening claim of this chapter, that some first person judgments are immune to error through misidentification relative to a use of the first person concept.

3.1 What is immunity to error through misidentification and why is it important?

First, some bookkeeping. An error of misidentification is the kind of mistake manifested by ascriptive judgments (of the form *a is F*) whenever a thinker goes wrong in identifying the object to which she is making an ascription. Suppose, for instance, that I see a white haired figure walking towards the lecture theatre, and on that basis form the judgment that professor X is on her way to class. As it turns out, I am mistaken; the person I see is not professor X but a student on the course. I am not mistaken about whether the person I see is on her way to class, I have simply gone wrong in my judgment about who that person is. I have made an error of misidentification.

Immunity to error through misidentification is a modal notion concerning the impossibility attaching to some judgments of mistakes of this kind. The corresponding positive modal notion relating to the possibility of errors of this kind is that of *vulnerability to error through misidentification*. I will treat both immunity and vulnerability to error through misidentification as features of judgments (rather than sentences, utterances, statements, assertions or propositions) relative to the grounds on which they are made. This is standard, but will be important later.[2] A pair of judgments with exactly similar contents can differ in their admittance of the possibility of errors of misidentification only if there is also a difference in the grounds on which they were formed.

Care must be taken to keep apart the notions of immunity and vulnerability to error through misidentification from what Lucy O'Brien has called *transparent immunity (vulnerability) to error through misidentification*.[3] The notions of

[2]But see (Shoemaker 1968), (Pryor 1999), and (Smith 2006) for some exceptions.

[3](O'Brien 2007, pp.212-13); O'Brien only discusses transparent *immunity* to error through misidentification, but there is no reason to suppose there is no corresponding phenomenon for vulnerability to error too.

immunity and vulnerability to error through misidentification directly concern the impossibility or possibility of misidentificatory errors given the epistemic conditions under which a judgment was formed. The notions of *transparent* immunity and vulnerability to error concern the accessibility of these facts to the judging subject. I mention this distinction only to note the danger of conflation; the interest of this chapter is exclusively in immunity and vulnerability to error through misidentification.

There is also another distinction to be noted and set aside. In his 1999 paper, *Immunity to Error Through Misidentification*, Jim Pryor draws a distinction between what he calls *immunity to* de re *misidentification*, and *immunity to which-object misidentification*.[4] The posited contrast tracks a distinction between two different kinds of error through misidentification. In the case of *de re* misidentification error, the judging subject misidentifies a particular person or object in the world for the pronounced object of her judgment. It is a mistake of this kind, for instance, that I was making in the above described scenario in which I formed a judgment about professor X: I misidentified the particular white haired figure I could see with professor X. A mistake of *which-object* misidentification, by contrast, occurs when a subject goes wrong in picking out the object to which the predicate applies in the first place. Upon hearing a loud tutting behind me, I might form the existential judgment that behind me there is someone who is getting impatient with my slow progress with the ticket machine. It is not difficult to imagine a development of this story in which I go wrong in identifying who the impatient queuer is — perhaps I glare angrily at the man directly behind me, when the tutter was, in fact, the woman behind him. This would be an error of which-object misidentification. The two different kinds of immunity, according to Pryor, correspond to the impossibility of committing these two different kinds of mistake. The discussion of the remainder of this chapter is intended to span both kinds of immunity.[5]

[4] See (Coliva 2006) for an attack on the viability of this distinction; but see (Prosser 2012), (Wright 2012), and (Campbell 2012) for defences of it.

[5] According to Pryor, wh-immunity to error through misidentification entails *de re*-immunity to error through misidentification, but not the other way around (see (Pryor 1999, p.286)). The explanation that follow in the next few sections have the structure of explanations of wh-immunity to error through misidentification, but if it is right that wh-immunity to error through

A number of contributors to the literature on immunity to error through misidentification — Annalisa Coliva, Campbell, O'Brien, Recanati, Crispin Wright and Simon Prosser, among others — have further distinguished varying *strengths* of immunity to misidentification errors.[6] This is a trend that can be traced back to Sydney Shoemaker, who draws a contrast between *absolute* and *circumstantial* immunity to error through misidentification — a difference, as we might now put it, between grounds-neutral and grounds-relative immunity.[7] As I have characterised the notion here immunity to error through misidentification does not come in degrees. It is always grounds-relative, and exists just whenever there is no metaphysical possibility of an error of misidentification of the kind described above.[8] I will not attempt engagement with each of the above supposed gradations in turn, but here is a very general comment about what I think these authors have got right — I agree that the significance of immunity to error through misidentification comes in different varieties. In the next section I will suggest that immunity to error through misidentification is susceptible to *explanations* of different kinds in different cases, and that this difference in turn gives rise to varying kinds of significance that should be accorded to the fact that a judgment is so immune. I think that these authors are right that immunity to error through misidentification is not a unified phenomenon.

These last remarks point in the direction of important questions of philosophical significance. Our judgments display patterns of immunity and vulnerability to a great many different sorts of error. Why is immunity to *this* kind of error important?

The notion of immunity to error through misidentification first emerged in Wittgenstein's *Blue and Brown books* (1958), in which he draws a distinction between two different uses of the word 'I': its use *as subject*, and its use *as object*. The latter, but not the former use, involves 'the recognition of a particular

misidentification entails *de re*-immunity to error through misidentification, this entails immunity of the second kind will hold too for those cases.

[6]See (Pryor 1999), (Coliva 2006) (Coliva 2012), (Campbell 2012), (O'Brien 2007), (Recanati 2012a), (Wright 2012) and (Prosser 2012)

[7](Shoemaker 1968)

[8]In this I am in agreement with Robert J. Howell that no judgments are absolutely immune to error through misidentification; see (Howell 2011)

person, and there is in these cases the possibility of an error'.[9] The notion of immunity to error through misidentification, then, is first introduced into the literature as significant because indicative of a special subjective perspective thinkers can hold towards themselves, an idea that continues to hold ground among many commentators of immunity to error through misidentification.[10]

The scope of the phenomenon of immunity to error through misidentification, however, is much wider than that of *first personal* immunity to error through misidentification; it also seems to hold for some perceptual demonstrative, pure indexical, mathematical, and descriptive name judgments, and has even been argued to hold — given the right sorts of gerrymandered conditions — for judgments relative to *any* kind of concept.[11] But it is unlikely that the significance of the immunity to error through misidentification of judgments containing descriptive nominal concepts, for instance, or perceptual demonstrative concepts, is likewise that it marks a subjective perspective on their referents. At the very least, then, more must be said to defend this supposed significance of first personal immunity.

The suggestion of this chapter is that the right place to start in asking questions about the significance of immunity to error through misidentification is with the explanation of its presence. After all, the fact that we are sometimes immune to such errors does not seem to be of very great interest in itself. The possibilities and impossibilities of error of this kind are, rather, markers of something much more interesting. What is philosophically important is whatever it is that *explains* the obtaining of immunity to error through misidentification for some first person judgments, and not for others.

Some authors have attempted to provide a homogeneous explanation for all instances of immunity to error through misidentification. Robert Howell and Cheryl Chen, for instance, both take the immunity to be a uniform corollary of the single-object dedication of the information channel grounding the judgment (and on that basis, question the importance of the notion). Crispin Wright argues that the immunity is always and exhaustively explained by the

[9](Wittgenstein 1958, pp.66-7)

[10]See, for instance, (Morgan 2015b),(Chen 2011), (Cassam 1997), (Hamilton 2009), (Recanati 2012a)

[11]The latter claim is in (Howell 2011)

non-inferential nature of the singular judgment. Simon Prosser takes all instances of immunity to error through misidentification to derive from a use of a singular term that is stipulated (rather than judged) to apply on the basis of the information grounding the judgment. Recanati explains immunity to error through misidentification in all cases by the absence of a singular component in the representational content of the experience on which the judgment is based, and goes some way to defend this account in the (perhaps slightly implausible) case of demonstrative judgments.[12]

Other authors — such as Andy Hamilton, Campbell and Peacocke — reject this assumption of homogeneity, as I will too in the next section.[13] This rejection ramifies: if the significance of immunity to error through misidentification supervenes on its explanation, and if there is no uniform explanation for all cases, then there will be no uniform significance attaching to the phenomenon of immunity to error through misidentification.

3.2 Three explanation-types

One principal dividing line between different explanations of immunity to error through misidentification, recognised early on by Evans, falls between those relativised to uses of purely descriptive concepts, and those relativised to non-descriptive concepts.[14]

A classic example of the former kind is due to Campbell:

> There is an easy way to generate judgments which are immune to error through misidentification. Suppose we introduce a descriptive name, "Frank", which has its reference fixed by "the inventor of the postmark." [...] [Y]ou form the judgment "Frank was the sole inventor of the postmark." You are fallible about this. [...] But there is a kind of mistake you cannot have made. It cannot be that you are right about there having been

[12](Howell 2011), (Chen 2011), (Wright 2012), (Prosser 2012), (Recanati 2012a).

[13](Hamilton 2009), (Campbell 1999b), (Peacocke 2008), (Peacocke 2012)

[14]Or, as he would put it, between information-based and non-information-based judgments; (Evans 1982, p.181); by 'descriptive concept' I mean a concept that refers, if it refers at all, to the unique satisfier of a description.

> a sole inventor of the postmark, but you are just wrong about which person it was. (Campbell 1999b, p.90)

It is plain to see how readily such cases may be generated, but what is the explanation of the immunity?

In cases of this kind, the immunity is explained by an agreement between the way in which the reference of the singular term or concept is fixed (in this case, by the description *the inventor of the postmark*) and the content of the ascriptive predicate (that of *being the sole inventor of the postmark*.) Given that the term 'Frank' is stipulated to refer to the unique satisfier of the description in question, a thinker cannot be correct in thinking that the corresponding descriptive predicate is correctly ascribable to *someone*, but wrong about whether or not it's correctly ascribable to Frank — that it correctly applies to the Frank is built into the very way in which the descriptive name functions to refer. This is Campbell's explanation too. He writes of the case:

> This suggests a connection between immunity to error through misidentification, and the way in which the reference of a singular term is fixed. It suggests that the way in which it happens, that there are judgments which are immune to error through misidentification, is that there are descriptive conditions on the reference of the singular term. So when the subject uses his grasp of the singular term to articulate a judgment in which the descriptive conditions are said to apply to the referent of the singular term, the result will be a judgment that cannot involve an error of identification. (Campbell 1999a, p. 90)[15]

I have submitted that the significance of a judgment's being immune to error through misidentification depends on the relevant explanation of that fact. What is the significance of the immunity to error through misidentification of judgments subject to this first kind of explanation? Arguably, not much — as Evans remarks of immunity of this kind in a footnote, '[it] is not very important; no one could plausibly claim that it constitutes the epistemological anchor-point of all our singular knowledge' (Evans 1982, p.181, n.52). One thing we

[15]Campbell's point is at the level of language, but so long as it is accepted that there can be purely descriptive concepts too, the point holds in thought as much as in language.

can say is that the immunity of this judgment is revealing of the descriptive condition under which the reference of the term has been fixed, or the descriptive mode of presentation under which the judgment's subject must be thinking of the referent, which can be extracted directly from the content of the judgment's predicative component. But this procedure is unlikely to unearth any great surprises — the immunity in cases of this kind has been manufactured by stipulation of the descriptive condition of reference determination, together with the production of a matching predicate ascription.

In its detail, the explanation just given will find no purchase beyond judgments containing terms or concepts whose reference is descriptively determined.[16] It nevertheless provides us with a kind of explanatory template for cases of other kinds, under which we should look to the nature of the agreement holding between the predicative and referential components of the judgment for an explanation of immunity. More specifically, we can put the template as follows:

> *Template*: Immunity to error through misidentification is explained by a matching between the predicative component of the judgment and the conditions on a successful use of the concept to which the immunity is relativised.

In this first kind of case — the case of immunity to error through misidentification that is relativised to the use of a descriptive concept — the content of the predication and the conditions on a successful use of the concept converge on the stipulative descriptive condition of reference, that 'Frank' refers to the sole inventor of the postmark. It is this description that both forms the content of the subject's predication, and determines reference to its object. This gives rise to the first of our three explanation-types:

> *Explanation-type 1*: Immunity to error through misidentification is explained in some cases by a descriptive condition that both forms the content of the predication and determines the reference of the use of the concept to which the immunity is relativised.

[16] This includes descriptive name concepts, as in the example just given, but also the mental counterpart of attributive uses of definite descriptions (if there are any).

A second kind of case involves the use of concepts whose reference is not determined by a purely descriptive condition. Take, for instance, the perceptual demonstrative judgment *that book is in front of me* made in the normal way on the basis of visual perception in normal lighting conditions etc. The use of the perceptual demonstrative concept *that* is determined to refer to the object to which the thinker is perceptually attending, and this judgment containing that concept is immune to error through misidentification — I could not be wrong in judging, on those grounds, that something is in front of me, but wrong about whether it is really the book that I see that is there. If I am wrong that *that* book is in front of me, then my visual experience leaves me with no surviving grounds to judge that anything is. The judgment is immune to error through misidentification relative to the use of the perceptual demonstrative concept.

We can again draw on the above template to explain this immunity, but we will need to fill it in a little differently. That template stated that immunity to error through misidentification is explained in some cases by a matching between the predicative component of the judgment and the conditions on a successful use of the concept. In this case, they do not match in virtue of the predicate's content and the mode of reference determination both being traceable to a particular descriptive condition. In this case, rather, the match is between the perceptual information that determines the reference of my use of the concept, and the (same) perceptual information grounding my predicate ascription of *being in front of me*. Both the determination of reference and the predication are jointly sustained by one and the same episode of perceptual contact with the book. Because of this, I could not be right about whether the ascription applies, but wrong about what it applies to.

To bring the point out, consider a second judgment I might form about the seen book, the judgment that *that book belongs to my sister*, made under normal conditions. There is no convergence here between the grounds underpinning the referential and the ascriptive components of the judgment. The reference of the demonstrative is determined by my perceptual relation with the book, whereas the ascription is based on my identification of the book I see as one that I know (through testimony, say) as belonging to my sister. Given this mismatch between the grounds of its referential and the ascriptive components,

this judgment is predictably vulnerable to error through misidentification. I might be correct in thinking that *some* book belongs to my sister (namely, the one she told me about), but wrong about whether it is the one I am now seeing.

On the back of this case, a second explanation-type comes into sight:

> *Explanation-type 2*: Immunity to error through misidentification is explained in some cases by the holding of a perceptual relation that both (partly) gives rise to the predication and determines the reference of the use of the concept to which the immunity is relativised.

In these first two kinds of explanation, the judgment's immunity to error through misidentification is explained by a matching between the predicative component of the judgment (either its content or its grounds) and the conditions of reference determination of the referring concept. In the first case, both come down to a single descriptive condition, while in the second, both are sustained by the same incoming perceptual information. An observation like this might lead to the impression that, as Campbell writes, 'there are deep connections between the way in which the reference of a [concept] is fixed and the immunity to error through misidentification of judgments containing that [concept]' (Campbell 2012, p.4)). Not all cases of immunity to error through misidentification, however, clearly fit this pattern.

Take Evans' example of the judgment *there is laughter here*, made on the basis of auditory perception in standard conditions with a normally functioning auditory system etc. (Evans 1982, p.189). The judgment is immune to error through misidentification. I cannot be correct on those grounds in thinking that there's laughter somewhere, but wrong about whether there's laughter *here*. The question now is whether this immunity can be explained as with the first two cases by appeal to a matching between the predicative component of the judgment and the conditions of reference determination. It seems not.

The reason for this pessimism is that the predication in this case is made on the basis of auditory perception, and its content is that *there is laughter*. If the explanation of immunity to error through misidentification is to appeal to a matching between the predication (either its grounds or content) and the conditions of reference determination for *here*, then we will have to say that either

here either functions as an auditory perceptual demonstrative concept, or that its reference is determined by a description such as 'the place at which there is laughter'. Neither of these options should be taken overly seriously.

To take the second first, it would be bizarre indeed for my use of the concept *here* to have its reference descriptively determined in this way. The reference of uses of descriptive concepts is determined via the properties of the object referred to — in this case, the suggestion is, the location's property of there being laughter there. But surely our *here*-thoughts do not depend on any particular properties of the location referred to (and much less on the property of there being laughter at that location!). Our uses of *here* refer to their location of use, no matter what properties might or might not be instantiated there. This, then, cannot be the way to go. The other option was to treat *here* as an auditory perceptual demonstrative. On this view, our uses of *here* refer to the place from which auditory input derives, perhaps alongside other forms of perceptual information. The problem with this option, again, is that weighs our *here*-thought down with too many additional conditions. Our uses of *here* refer to their location of use, no matter what perceptual input might or might not be derived from it. It seems, then, that the explanation of the immunity to error through misidentification in this case cannot draw on a matching of the right kind between the predicative component of the judgment and facts about the mode of reference determination — the example is a bad fit for either explanation-type 1 or 2.

There is, however, another way of filling out the template that remains available. At the very heart of this dissertation is the idea that there is more to singular thought than referential success; there are referential conditions, but there are also conditions of comprehension. Explanation types 1 and 2 both drew on a match between the predicative component of the judgment and facts about the conditions of reference, but another way to go is to draw on a match between the predicative component of the judgment and facts about the conditions of comprehension. It is, I suggest, an explanation of this latter kind that is at play in the present case of the judgment *there is laughter here*.

How would the explanation go? Reference determination for *here*-thought is given by the rule that uses of *here* always refer to their locations of use. In

Chapter 2, we saw that this rule of reference determination generates certain epistemic conditions on concept-grasp — in order to grasp one's own use of *here*, one must be epistemically related to the location one is thinking of in a way that is constrained by the same overarching condition of location-identity that also constrains the reference of one's thought. One must, that is to say, be related to the relevant location through essentially location-reflexive epistemic relations. One likely-looking candidate for such an epistemic relation is auditory perception; auditory contact with a location arguably depends on the identity of the location at which the sounds are heard, and the location from which they immediately emanate. If this is right, then auditorily perceiving one's environment is one way to be in the right epistemic state to grasp one's own use of the concept *here*.

There are ways of resisting the claim that audition is an essentially location-reflexive epistemic relation. It might be objected, for instance, that I can perfectly well hear my mother's voice from thousands of miles away. The invention of Skype, phones, walkie-talkies etc. has extended our faculty of audition across space, so that it is no longer an essentially location-reflexive form of knowledge about a location; it's a way that I have of sometimes knowing about how things are in my current location, but it is also a way that I have of knowing how things are elsewhere. Indeed, the point does not even need appeal to such mechanisms. Imagine a wolf howling to the moon from a distant mountain top. There are, of course, unsettled questions the referential radius of my use of *here*. But suppose that we position the wolf on a mountain top so far away that it plausibly escapes the spread of my reference. Aren't these cases in which audition is used to find out about how things are in locations other than the one at which the epistemic agent is located?

It is crucial here that we distinguish between the sounds that are heard and their sources. What these cases show is that the *sources* of sounds need not be in the listening subject's immediate spatial environment — indeed, cases of recordings show that they need not be in her immediate temporal environment either. The sounds themselves however — the sound of my mother's voice, the howling of the wolf — are the objects of my audition. And it would seem hard to deny that *these* must be at the place where the subject is listening. (It is rea-

sonable to assume that my mother has spoken today. But given that the sound of her voice is miles away, I have not heard her). So long as we understand sounds to be the proper objects of audition, then, the claim that audition is a location-reflexive epistemic faculty remains on solid grounds.[17]

Auditory perception is playing double duty with respect to the subject's judgment that *there is laughter here*. It is both grounding the subject's predication of laughter, and it is also providing her with a way of grasping her own use of the concept *here*; the very same perceptual information both gives rise to the predication, and allows the thinker to comprehendingly think a *here*-thought at all. This explains why the judgment is immune to error through misidentification. The thinker could not be correct in judging, with full comprehension and on the grounds described, that there is laughter *somewhere*, but not that there is laughter *here*, because the very auditory information that would underpin the existential judgment of laughter at one and the same time also underpins the thinker's use of the concept *here* to comprehendingly refer to the place she is in. With this case in view, then, we can offer a third explanation type:

> *Explanation-type 3*: Immunity to error through misidentification is explained in some cases by the holding of a perceptual relation that both (partly) gives rise to the predication and form part of the conditions on the thinker grasping her own use of the concept to which the immunity is relativised.

Cases of immunity to error through misidentification that are subject to this third kind of explanation are not revealing of the mode of reference determination for the relevant singular concept, but rather tell us something about what it takes for a thinker to be in a position to comprehendingly use that concept in the first place. The significance of the fact that an auditory *here*-judgment is immune to error through misidentification is that it tells us that audition is at least one of the forms of acquaintance that can underwrite comprehending uses of the *here* concept.

[17] It is worth pointing out that to the extent that one resists the claim that audition is a location-reflexive epistemic channel, one will also have to deny that cases like *there's laughter here* will be immune to error through misidentification; the explanandum stands and falls with the proposed explanans.

Earlier, I said that immunity to error through misidentification is a property of judgments — rather than one of, say, of propositions or sentences. We can now see why this is important. Unlike propositions or sentences, judgments are conscious mental episodes. Conscious mental episodes, we saw in chapter 2, characteristically involve comprehension on the part of the thinker; it was difficult to find a pure case at the level of thought in which the conditions of reference and the conditions on comprehension came apart. In ordinary circumstances, when I judge something, I grasp what it is that I am judging. This means that in ordinary circumstances, explanations of type 3 will always be available wherever there is a match between the grounds of predication and the grounds of such concept-grasp. To compare: if immunity to error through misidentification was merely a property of sentences, then it would be less clear how widespread this kind of explanation would be. *Sock-drawer* of chapter 2 showed the relative availability of linguistic cases of reference without comprehension. So if immunity to error through misidentification were a property of sentences, more work would need to be done in each case to demonstrate the presence of comprehension before an explanation of type 3 could be applied.

Before turning back with these explanation-types in hand to the asymmetry challenge for first person thought, it is worth taking a moment to consider the way in which the explanations just given for the three kinds of cases discussed in this section fit together with a somewhat more traditional explanation of immunity to error. This is an explanation given in terms of the identification-freedom of the judgment's justificational, or perhaps presuppositional structure — if there is no identification in the judgment's grounds, the basic idea is, there is no possibility of a *mis*identification.

The lineage of this simple and popular approach is traceable to Evans, who explains that,

> When knowledge of the truth of a singular proposition $\ulcorner a$ is $F\urcorner$, can be seen as the result of knowledge of the truth of a pair of propositions, $\ulcorner b$ is $F\urcorner$ (for some distinct Idea b) and $\ulcorner b = a\urcorner$, I shall say that the knowledge is *identification-dependent*.' (Evans 1982, p.180).

Immunity to error through misidentification, Evans goes on to explain, is a

corollary of the absence of such identification-dependence — or, as he puts it, a corollary of *identification-freedom*.[18] The approach is elegant in its simplicity, and survives under a range of precisifications in the current literature as a dominant explanatory approach to immunity to error through misidentification. As Wright writes of it, it 'liberates us from any need for metaphysical or semantic extravagance in trying to account for the phenomenon. It is a pleasingly deflationary account.' (Wright 2012, p. 255)[19]

The important point to note here is that the three explanation-types offered in this section are not in competition with explanations of this kind. Identification-freedom, rather, and the explanation-types 1, 2 and 3 are simply lodged at different levels of explanation; identification-freedom at a higher level of generality, and the more specific explanations at a level lower down. Indeed, the lower level explanations can themselves be seen to *explain* the identification-freedom of judgments with immunity to error through misidentification. Take, for instance, an explanation of type 2: that the perceptual grounds of the predicate ascription coincide with the grounds of the concept's reference determination can itself be taken to explain the judgment's structural identification-freedom — if the very perceptual input that fixes the singular concept's reference also provides immediate warrant for the predicate's ascription, then there will be no need for a mediating identification in ascribing the predicate to the concept's referent. Parallel moves are likewise available for explanations of types 1 and 3.

In a manner of speaking, then, we can help ourselves to the simplicity of the identification-freedom explanation at a higher explanatory level, while also reaping finer-grained accounts of the significance of the immunity of these judgments from the finer-grained explanations at the lower level.

[18] p.183

[19] See also (Smith 2006), (Coliva 2006), (Peacocke 2014) for some other proponents of this approach.

3.3 The asymmetry challenge

At the start of this chapter I said there was a prima facie explanatory challenge facing accounts of first person thought, to explain why it is that some first person judgments are, and some aren't immune to error through misidentification relative to a use of the first person concept. That challenge, I added, is brought into sharper focus for accounts under which the mode of reference determination for the first person concept is given by the rule that its uses always refer to their thinkers. That is because those accounts cannot straightforwardly appeal to a matching between the mode of reference determination and the predicative element of the judgment (either its grounds or its content) to explain the immunity when it arises. So we are left facing the question what can.

The unsuitability of explanation by appeal to mode of reference determination in the case of *I* runs parallel to the *here*-case we've already considered. Given that audition-based judgments are immune to error through misidentification relative to uses of *here*, we saw that an explanation of type 1 or 2 — an explanation that relies on the mode of reference determination for *here* — would force us into accepting one of two wildly unattractive views about the reference determination for *here*: either that it is fixed by a description like 'the place at which sounds are heard', or that it should be treated as an auditory demonstrative. The problem in the case of the first person is analogous to these problems for *here*. First person judgment-types with immunity to error through misidentification arguably include at least introspection, bodily awareness (of certain kinds) and memory judgments. To say that the explanation of these cases of immunity are of type 1 would be to say that uses of *I* have their reference fixed by something like the description, 'the unique source of this introspective, bodily awareness and memory input'; to appeal to explanations of type 2 would be to submit to the demonstrative model of the reference determination of *I*. Either way, commitment could not be maintained to the view that the reference of *I* is exhaustively determined by the rule that its uses refer to their thinkers, or producers.

The momentum of this asymmetry challenge for self-reference rule accounts of first person thought is clearly articulated by Campbell:

> Now, so long as we take it that the reference of the first person is fixed by the token-reflexive rule: 'Any token of "I" refers to whoever produced it', the thing seems completely baffling. There is no way of explaining why there is this apparent asymmetry [...], so long as that rule is all we can appeal to in giving the explanation' (Campbell 2012, p.13).[20]

Campbell's reaction is extreme. In the case of the first person — and unlike all other terms or concepts — he urges us to simply allow that the immunity to error through misidentification of some of our first person judgments, and the vulnerability to such error of others, is just an aspect of the patterns of use for the first person, which is subject to its own 'bedrock, normative correctness'. (Campbell 2012, p.21)

I think Campbell is right that this challenge seems completely baffling so long as we restrict the materials available in explanations of immunity to error through misidentification to facts about the mode of reference determination (along with facts about predicative grounds of the judgment). The point of the last section, however, was to push for a deepening of the resources available for use in explanations of immunity to error through misidentification. It is true that there is a central range of cases in which the presence of immunity to error through misidentification is explained by appeal to the concept's mode of reference determination, as in explanation-types 1 and 2. But, I argued, there are also other cases in which the explanation comes not from the conditions of reference determination, but from the conditions on understanding. These are the cases falling under explanation-type 3.

The first person case, I suggest, is one of them. To make the point, take as a representative example of a first person judgment with immunity to error through misidentification the thought *I am thinking of a canary* made on the basis of introspection. The immunity to error through misidentification of judgments

[20] If the first person concept has its reference determined by the rule of subject-reflexive reference, then it will also be possible to engineer first personal judgments that are immune to error through misidentification in a way that *is* explained by appeal to that rule of reference determination, in the style of the 'Frank' example above: e.g. the judgment that *I am the producer of this thought*. These, however, are artificially constructed cases and do not form part of the asymmetry challenge, which is to explain the naturally arising patterns of immunity and vulnerability to error through misidentification relative to uses of the first person concept.

of this kind will be defended in Chapter 4, but for now it is enough to note the strong plausibility of the verdict that I could not know on those introspective grounds that *somebody* is thinking of a canary, but be wrong about whether that person is me — indeed, some commentators take the immunity of such judgments to be an inviolable commitment of adequate accounts of immunity to error through misidentification.[21] The question is, how it this immunity to be explained? The suggestion is, by a type 3 explanation.

Type 3 explanations are explanations by a matching between the grounds of predication and the conditions on a subject's grasping of her own use of the concept. The conditions on grasping one's own use of the first person concept, we saw in Chapter 2, are that the subject must be epistemically self-related in ways that align with the reference of her thought. She must be self-acquainted, that is to say, in ways that are essentially subject-reflexive forms of self-knowledge — forms of self-knowledge that depend *de jure* on the identity of the knower and the known. One such source of self-knowledge is introspection; the way in which I come to know about my mental state through introspection is one that depends on my identity with the person whose mental state is known about. This explains the immunity to error through misidentification of my introspective judgment that *I am thinking of a canary* as follows: I could not be right on those grounds in thinking that someone is thinking of a canary but wrong that it's me, because one and the same act of introspection grounds both my capacity to comprehendingly use the concept *I* to refer to myself, and the predicative component of the judgment. Introspection, to put it another way, plays a double role with respect to my judgment. It both allows me to grasp my use of the concept in forming the judgment, and it gives rise to my predication (to the person so thought about) of thinking of a canary.

There are also, of course, first person judgments that are vulnerable to errors of this kind making up the other side of the asymmetry challenge. Take, for example, the judgment *I am looking tired*, formed on the basis of catching sight of a worn looking figure in a backbar mirror. If there is a type 3 explanation avail-

[21] See (Langland-Hassan 2015) and (Smith 2006); the immunity to error through misidentification of introspective first person judgments is nearly universally accepted, but see (Campbell 1999b) for an example of dissent.

able to explain cases of first person judgments that *do* display immunity to error through misidentification, we should expect to find an absence of such an explanation here — which is just what we do find. The grounds of predication in this case is an episode of visual perception involving mirrors. This is, of course, a form of self-knowledge of a kind; it is a way that I sometimes have of finding out how things are with myself. It is not, however, a form of self-knowledge that plays the kind of enabling role for first person thought described in Chapter two. That is because it is not an essentially subject-reflexive form of self-knowledge, or a form of self-knowledge that aligns with the referential conditions of my thought. There is no match here, then, between the predicational grounds and the comprehension conditions on my thought — and so, according to the present proposal, no immunity to error through misidentification. The asymmetry challenge, then, is no longer quite so baffling once we bring explanations of type 3 into play.

Once again, it should be noted that there is nothing in this explanation of the patterns of first personal immunity and vulnerability in terms of comprehension conditions that competes with a higher level explanation given in terms of the identification-freedom or -dependence of a first person judgment's grounds or presuppositional structure. Once again, rather, the fact that there is a matching of the conditions of comprehension and the grounds of predication can be taken to itself *explain* such identification-freedom; if one and the same introspective episode grounds my comprehending use of *I* to refer to myself as well as giving rise to the predication that I ascribe to myself (so thought about), it is no wonder that I need draw on no further identificatory information about the situation in order to judge that the predicate applies to me.

3.4 The significance of immunity to error through misidentification relative to uses of *I*

Earlier I suggested that the best place to start looking for the significance of a judgment's being immune to error through misidentification is with the explanation of its presence. I have now offered an explanation of the presence of immunity to error through misidentification for a range of first person judgments

— namely, that it reflects a matching between the conditions of comprehension on first person thought and the grounds of certain kinds of predication. What is the significance of this explanation of immunity to error through misidentification?

Just as we could work backwards in the descriptive name case from the presence of immunity to error through misidentification to the descriptive condition fixing the name's reference, the significance of this explanation is that we can likewise work backwards from verdicts of immunity and vulnerability to error through misidentification in the first person case to discover the specific forms of self-knowledge underpinning our comprehending uses of *I*.

The discussion of Chapter 2 delivered the resources to say that the forms of self-knowledge that enable grasp of uses of the first person concept are those that meet *mutual constraint* (p.46); they are those that are constrained by the same overarching condition constraining the thought's reference. To be precise, they will be all and only our *de jure* subject-reflexive forms of self-knowledge. To say this, however, is to say very little about what those forms of knowledge will actually look like. In that chapter I suggested that they might plausibly include such faculties as internal bodily awareness and introspection, among others. If this is right, however, then it is a contingent and empirical fact about the kinds of creatures that we are — organisms differently built might well have very different kinds of channels of self-knowledge meeting the same constraint.

As a matter of empirical investigation, then, there is a pressing question about what the markers of these forms of self-knowledge are that will allow for their recognition. The present suggestion is that one such marker will be the immunity to error through misidentification of first person judgments made on their basis. That is because the role of these forms of self-knowledge in enabling comprehension of our own first person thought is what *explains* the immunity to error through misidentification of first person judgments made on their basis. If we unpack this line of thought in the opposite direction, this means that the immunity to error through misidentification can in turn be taken as a reliable indicator of a judgment's grounds being one of the forms of self-knowledge that enable the thinker's grasp of the use of the first person concept. To discover which are the forms of self-knowledge that play this fundamental role in our

grasp of our first person thoughts, then, we must look to the grounds of first person judgments that are immune to error through misidentification.[22] This is the significance of first personal immunity to error through misidentification.

3.5 Conclusion

The aim of this chapter was to meet the asymmetry challenge facing self-reference rule accounts of first person thought. Given that the first person concept is governed by the rule that uses of the concept refer to their thinkers, the question was, how are we to explain the asymmetry between those first person judgments that are and those that aren't immune to error through misidentification? In response, I advocated for an increase in the resources at our disposal in explanations of immunity to error through misidentification. Although it is right that such immunity is sometimes explained by facts about the concept's mode of reference determination — a strategy that is seemingly unavailable in the case of *I* — it is elsewhere explained by appeal to facts of another kind, by facts about the conditions of comprehension rather than those about the conditions of reference. It is an explanation of this latter kind, I argued, that applies to the case of the first person concept.

In the last section, I argued that this explanation makes the immunity to error through misidentification of first person judgments important to us in a very particular way: it provides us with a way to recognise the forms of self-consciousness that underpin our comprehending uses of the first person concept. This is a task that is now due.

[22]Strictly speaking, immunity to error through misidentification of judgments made on its basis is necessary but not sufficient for a form of self-consciousness to be cast in this role – i.e. if judgments formed on its basis are *not* so immune then it *cannot* be one of the special forms of self-consciousness, but if it is then it's always possible that there could be some other explanation of it than that it plays this role (I don't take myself to have exhausted all possible explanations of immunity to error through misidentification with the three explanation-types of §2). Given that it does not seem as if the alternative explanation could be to do with reference fixing, though, not clear what it would be. At the very least, then, we can say that the immunity to error through misidentification of judgments made on the basis of a given form of self-consciousness is a *good guide* to its playing this role.

Chapter 4

Which forms of mental awareness?

The preceding chapters have charted an important role for a privileged cluster of forms of self-consciousness. We began, in chapter 1, with the idea that there are some forms of self-awareness that enter into a characterisation of the conceptual role for *I*, while others are left out. The challenge was to say how we could derive such facts about the canonical pattern of use that we make of the first person concept from the rule governing the concept's reference that uses of *I* refer to their thinkers. The answer, in chapter 2, was that this special group of forms of self-awareness are those that underpin our capacity for fully comprehending first person thought. It is their playing this central role in the conditions of *comprehension* on first person thought that explains this aspect of the conceptual role for *I*, rather than any role played directly in the conditions of reference. Given that the constraining condition on such forms of self-consciousness can be derived from the rule of reference determination, this provided an answer to the challenge of chapter 1.

This special cluster of forms of self-consciousness, then, plays a fundamentally important role in a full account of first person thought: they enable us to comprehend our own episodes of first person thinking. I further argued in chapter 3 that these forms of self-consciousness are those that give rise to self-ascriptive judgments with immunity to error through misidentification relative to a use of the first person concept, and so that such immunity is a useful indicator of the forms of self-awareness that play this important role.

Up to this point this role has been carved out for this privileged cluster of

forms of self-consciousness at a certain remove from details about what they will actually look like. Rather, these forms of self-consciousness have largely been characterised by the general condition of essential subject-reflexivity, and the discussion has been focussed on questions about what work forms of self-consciousness meeting this condition — whatever they turn out to be — could do. But this condition of essential subject-reflexivity itself says nothing about which are the forms of self-consciousness that will meet it; creatures differently configured will have essentially subject-reflexive forms of self-consciousness of different kinds. This chapter and the next finally asks, what are these forms of self-consciousness like for creatures like us?

What follows is two chapters of two parts each. The aim of each part is to establish the immunity to error through misidentification of first person judgments based on a different faculty of self-consciousness: introspection and episodic memory in this chapter, and internal bodily awareness and multimodal bodily awareness in the next. For the most part, for each of these four faculties it is initially plausible to think that self-ascriptive judgments made on their basis will be immune to error through misidentification relative to a use of the first person concept. My stance in each of these sections, then, is defensive. I attempt to deflect various challenges to the assumed immunity to error through misidentification of the first personal judgment-kind in question. These challenges come in two forms, empirical and conceptual. Challenges of the former sort draw on real-life cases — typically, pathological — constituting apparent counterexamples to the claim that the judgment-kind in question is immune to error through misidentification relative to uses of the first person concept. Those of the second sort appeal to the mere conceivability of such cases to make the same point. I consider challenges of either kind as appropriate.

Altogether in these two chapters I consider a fair range of such cases, and the arguments in defence of the immunity to error through misidentification of the associated judgments are peculiar in each case to the details of the faculty involved and of the cases considered. To keep the dialectic manageable it will perhaps be helpful to set out a brief anticipatory inventory of the cases and responses here for reference.

· (Chapter 4) I consider two empirical counterexamples to the immunity to error through misidentification of introspective judgments: the delusion of thought insertion and the rare phenomenon of conjoined twins joined at the brain. In response to the first, I make a familiar argument that the case has the wrong formal structure to be a counterexample to the immunity to error through misidentification of introspective judgments. In response to the second, I argue that absorption into one's mental life is sufficient for an introspected mental state to count as one's own, so the case is not yet one in which a self-ascription amounts to an error of misidentification.

· (Chapter 4) For the faculty of memory I consider the purported conceptual counterexample of quasi-memory. I argue that the construal of quasi-memory as a threat to the immunity to error through misidentification of memory judgments relies on an outdated psychological model of memory.

· (Chapter 5) In the case of internal bodily awareness I consider both a conceptual case — a case of redirected wires — and, briefly, an empirical one — somatoparaphrenia. The response to the second is that, analogous to the thought insertion cases, it has the wrong form to constitute a counterexample to the immunity to error through misidentification of internal bodily awareness judgments. The response to the first is that redirected wire cases are typically underdescribed. Once the details are filled in the cases become either so weak as to be benign or they collapse into internal incoherence. Somatoparaphrenia, like thought insertion, is a pathology that issues in judgments of the wrong shape to count as a counterexample.

· (Chapter 5) Finally, I consider three empirical cases as potential counterexamples to the immunity to error through misidentification of multimodal bodily awareness, in the form of three clinically induced illusions: the rubber hand illusion, the body transfer illusion and the nose displacement illusion. My response to these cases comes in two stages. First, I argue that we have reason to reject an inferential — and so identity-

involving — reading of the process through which our multimodal experiences of our bodies is produced. Second, I offer an alternative understanding of the mistake being made in these three illusions.

Unlike the foregoing three chapters, the task of this pair of chapters is chiefly empirical in nature — in each section it is asked whether an actual epistemic faculty of ours meets a particular modal criterion, namely, that of giving rise to judgments with immunity to error through misidentification. As such is vulnerable to pitfalls of different kinds. Putting the task this way, for example, already involves something of a fiction. As we go about in our daily knowledge-involving activities, our epistemic sensitivity to our environments and to ourselves does not come neatly packaged in separate and clearly labelled epistemic faculties; much more familiar, rather, is an incoming buzz of unsorted information. Indeed, results in recent work in crossmodal perception are unequivocal that we are spectacularly bad at recognising which sense modalities are responsible for which perceptual judgments.[1] Even a mere statement of the chapters' task, then, already imports theoretical assumptions that are liable to complicate interpretations of the purported counterexamples. This is not to agitate for scepticism about the existence of individual epistemic faculties, only to urge caution with regards to this methodological difficulty. To treat the different forms of awareness as separately as I do here is no doubt something of an idealisation, but I think it is one that is harmless for our purposes. My aim is to identify the forms of awareness that enable grasp of our uses of the first person concept. Those forms can play that role, even if we rarely actually experience the exercise of those forms of awareness independently from the exercise of others.

This is not the only methodological difficulty these chapters face. When working to rebut counterexamples incompleteness is unavoidable. There might always be new counterexamples, or new ways of dealing with the current ones emerging from fresh empirical data. These chapters are just a start in an ongoing empirical project to understand the nature of self-knowledge for

[1] For just a few examples, see (MA Heller 1999), (GF Misceo 1999), (T Seizova-Cajic 2011), (T Seizova-Cajic 2007), (Claudia Lunghi 2014), (Y Jiang 2013), (V Occelli 2013), (Nadia Bolognini 2012), (Shimada 1990).

the kinds of creatures we are.

4.1 Introspection

Introspection is the non-observational, non-inferential epistemic faculty by which we come to know about our occurrent mental states, events and processes. There is an established history of treating introspection as an especially epistemically secure form of self-knowledge; claims about the infallibility, incorrigibility, self-intimation, luminosity, and indubitability of our introspective access to our mental lives have come in and out of philosophical fashion in repeated attempts to capture the seemingly distinctive epistemic privilege of introspection. We can add to this list the claim — sometimes given as partly definitional of the faculty of introspection — that introspection allows us epistemic access to our own mental life and none other.[2]

The claim that introspection only provides a subject access to her *own* mental states (events etc) feeds directly into a corollary claim about the immunity to error through misidentification of introspection-based first person judgments:

> *Introspective immunity:* I cannot be wrong, in forming a judgment about an occurrent mental state, event or process of the form *I am F* (or *am f-ing*) on the basis of introspection solely in virtue of a misidentificatory mistake about whether or not it is *me* that I know to be *F* (*F-ing)*on those grounds.

Even where the claim of single subject access is not treated as a stipulative feature of introspection, *introspective immunity* has rarely been challenged. Indeed, in a move that has been cited with approval by others, Joel Smith presents the immunity to error through misidentification of such judgments as a non-negotiable constraint on precisifications of immunity to error through misidentification.[3] He explains:

> [A]n account of IEM should, at the very least, capture those self-ascriptions that are agreed by all to be central to our conception of ourselves as self-conscious subjects. Specifically, an account of IEM should imply that the

[2]See (Schwitzgebel 2014) for a record of some of these epistemic privileges that have been historically associated with introspection.

[3](Langland-Hassan 2015, p. 3)

> self-ascription of occurrent mental episodes (e.g. 'I have a headache') are IEM. (Smith 2006, pp. 7–8)

In a similar vein, it will be remembered that I helped myself in chapter 3 to introspection-based judgments as a relatively uncontroversial example of a first person judgment-kind with immunity to error through misidentification. Even if relatively uncontroversial, however, there have nevertheless been at least two empirical challenges to *introspective immunity* in the literature: the schizophrenic delusion of thought insertion, and cases of craniopagus conjoined twins. Neither, as it turns out, will be enough to undermine *introspective immunity*.

4.1.1 Thought insertion

Thought insertion is the delusion — currently treated as a key diagnostic symptom for schizophrenia — in which sufferers believe some of their thoughts to have been placed into their stream of consciousness by somebody or something other than themselves. This might seem to pose a direct counterexample to the claim that errors of misidentification are impossible on the basis of introspection. Perhaps the most influential expression of this challenge to the immunity to error through misidentification of our introspective judgments comes from Campbell, who writes:

> A patient who supposes that thoughts have been inserted into his mind by someone else is right about which thoughts they are, but wrong about whose thoughts they are. So thought insertion seems to be a counterexample to the thesis that present-tense introspectively based reports of psychological state cannot involve errors of identification.(Campbell 1999b, p. 610)

Sufferers of thought insertion, the idea is, form judgments on the basis of introspection about an occurrent psychological state in which she is correct in judging that *someone* is undergoing that state, but mistaken through a misidentification as to *who* it is. The subject has misidentified the thing that is undergoing the mental occurrence (that is, herself) with another object (the thought's attributed source.)

The nature of the threat posed by this delusion can be seen as taking the following form. That these cases involve an identification component in the formative grounds of the judgment might seem to demonstrate the involvement of an identification (perhaps between oneself and 'the thinker of this thought') in *all* introspective judgments. Cases of thought insertion seem to force a review of our understanding of the process by which our typical introspection-based self-ascriptions are formed, because the possibility of a misidentification error in thought insertion cases demonstrates our introspection-based judgments to involve an identification-component — even if that identification never normally goes wrong in healthy subjects. As Shaun Gallagher puts the challenge, to allow that thought insertion cases involve errors of identification, 'would involve admitting [...] that first-person awareness [...] does involve an identification, that schizophrenics get it wrong and that normal subjects get it right.' (Gallagher 2000, p. 208) If this is right, then our introspection-based judgments cannot be immune to error through misidentification.

To see why thought insertion is not a counterexample to *introspective immunity*, it will be helpful to first set out some terminology developed elsewhere by Frederique de Vignemont. De Vignemont draws a helpful distinction between what she calls *false-negative* and *false-positive* errors, a distinction she presents as follows:

> There is a *false negative* if one does not self-ascribe properties that are instantiated by [oneself]. False-negative errors have to be contrasted from false-positive errors. There is a *false positive* if one self-ascribes properties that are instantiated by another.

She goes on to explain that '[t]he hypothesis of [immunity to error through misidentification] clearly concerns false positives.' (de Vignemont 2012, [229)

This should be a point of agreement among all theorists of immunity to error through misidentification — to say that a particular class of self-ascriptive judgments, made on certain grounds, are immune to error through misidentification is to say that it is impossible to form such a first personal judgment on those grounds while being mistaken through a misidentification in taking oneself to be the object of one's judgment. In de Vignemont's terms, then, it

is a claim about the impossibility, under those conditions, of a false-positive error. *Introspective immunity* is just such a claim about the impossibility of introspection-based false positives errors.

It is only a small step from here to the recognition of compatibility of *introspective immunity* with the empirical phenomenon of thought insertion, for it is clear that thought insertion demonstrates not the possibility of introspection-based false positive errors, but of introspection-based false negatives. On the basis of introspection, these subjects fail to self-ascribe properties that they are in fact instantiating. Demonstrating the possibility of introspection-based false negatives, however, does nothing to threaten the impossibility of introspection-based false positives. Taking the false belief involved a case of thought insertion to be in error of misidentification is perfectly consistent with the claim that our introspection-based self-ascriptions of psychological properties are immune to error through misidentification. There is no threat here to *introspective immunity*.[4]

That concludes the defence of the immunity to error through misidentification of introspective judgments against thought insertion cases. Before moving on to the next purported counterexample, however, I want to touch on a potential objection of another kind. There might, the objection goes, still be call to think that even if the delusion is no threat to *introspective immunity*, it nevertheless causes problems for the proposed account of first person thought. The role so far blocked out in this dissertation for essentially reflexive forms of self-knowledge is one of providing grounding for the thinker's very capacity for comprehending first person thought, in the way described in chapter 2. One of the necessary markers of the forms of self-knowledge playing this role is that judgments formed on their basis will be immune to error through misidentification. Thought insertion cases do nothing to show that introspection lacks that marker. What they might show, though, is that introspection is not really fit to play this important role in underpinning our first person thoughts, because it just isn't the kind of self-knowledge that delivers up a subject to herself in a reliable enough way. To put the challenge another way, perhaps demonstrations of the possibility of false negative errors pose a threat of their own: perhaps

[4]See (Coliva 2002, p. 30) and (Langland-Hassan 2015, pp. 5–6) for similar arguments.

they show that the faculty in question is not tightly enough enmeshed with our first person ways of thinking to play the grounding role for first person thought that the proposed account has them playing.

This challenge is, I think, not much more than surface deep, but it is worth making explicit why this is. I can think of two things to say to this effect. First, the reasons we could cite for which a thinker might not exploit an essentially reflexive form of self-knowledge to form a first person judgment are as many and varied as our creativity capacity for story-telling will allow. Perhaps the subject is hit over the head at the crucial moment, or is in the grip of self-deception of a kind that prevents her from forming that particular sort of judgment. Perhaps there are repressive psychological mechanisms at work associated with post traumatic stress disorder, or maybe she just gets distracted before she gets around to forming the judgment. The point here is that the presence of a false negative error is of very little interest in itself; it tells us nothing yet about what happens when the subject *does* go on to form a first person judgment on the relevant basis. This is a quite general point about the significance of demonstrating the possibility of false negative errors in the context of the present discussion.

One might still object, though, that even if thought insertion cases are of no special interest on the strength of their basic form alone, they — unlike the stories told in the last paragraph — *do* show us something worrying about what introspection is like, that should lead us to question whether it can play the role this dissertation has it playing. The constraint on the possibilities for thought that is generated by being hit over the head is one that is one that comes from outside the subject and her mental processes. We don't think it shows us anything interesting about the cognitive processes involved in introspection because it is, as we might put is, an externally derived problem of damage to hardware. Although the other cases mentioned (lack of attention, or repression associated with self-deception or post traumatic stress disorder) involve causal influences or suppression mechanisms that are obviously not external to the subject's mental activities, they *do* seem to derive from something external to the introspective process itself. Like being hit over the head, these cases involve an *interruption* or *overriding* of what otherwise might have been the formation of a self-ascriptive introspective judgment. The phenomenon of

thought insertion doesn't seem to be like this. The kind of false negative error manifested by thought insertion judgments emerges from something within the experience of introspection itself. Unlike being hit over the head, this kind of disruption does seem to tell us something about what introspection is like. Isn't this a worry in itself?

The second thing to say — that speaks to this worry directly — is that we simply have no evidence that the psychological disruption that gives rise to reports of thought insertion is one that pertains to the phenomenology of experience itself, rather than a post-experiential addition or distortion or rationalisation of the original introspective state. That these subjects are ultimately unable to exploit this essentially reflexive form of self-knowledge does not show that basic exercises of introspection are not apt for immediate response with a first person thought; it only shows that this aptness must be screened off in some way for these thinkers. There is no reason to think, in other words, that thought insertion does not belong alongside the other stories of three paragraphs ago, as yet another way in which the transition from introspection to self-ascriptive introspective judgment might be interrupted or overridden.[5]

4.1.2 Craniopagus twins

In 'Introspective Misidentification' (2014), Peter Langland-Hassan discusses an empirical challenge to *introspective immunity* of another kind. He considers the case of Krista and Tatiana Hogan, a pair of twins conjoined at the head and brain. More specifically, they are connected by a 'bridge' between their respective thalami, a part of the brain thought to be responsible for the relay of information from subcortical areas to parts of the cerebral cortex. Other than this, the twins are neurophysiologically unremarkable; each has a fully intact left and right hemisphere connected by a corpus callosom, all of which is contained entirely in her own head.

The interest of the Hogan twins with respect to *introspective immunity* comes from the unusual capacities apparently generated by their unique physiology.

[5]Similar worries might arise for structurally analogous false-negative cases of other kinds, like the case of somatoparaphrenia that will be discussed in §5.1.2. Similar responses apply (*mutatis mutandis*) to those given here.

As Langland-Hassan explains:

> With her eyes covered by her mother's hand, Krista seems able to report on what kind of object (a toy pony) has been raised before Tatiana's eyes; facing the opposite direction, Tatiana knows when (and where) Krista is being tickled. Each twin seems to know what the other is seeing or feeling, and perhaps even thinking, in a way others cannot. Each seems to know these things through introspection. (Langland-Hassan 2015, p. 7)

By making use of what seems to be their ordinary faculty of introspection, these twins have apparent access to the visual and sensory experiences, thoughts, and emotions of the other. Langland-Hassan uses their case to argue that introspection-based errors of misidentification are logically (and perhaps nomologically) possible. Against Langland-Hassan, I will argue in this section that their case shows no such thing.

Let's suppose it to be true that each twin has introspection-involving access to the mental states, events and processes of the other (Langland-Hassan is perfectly upfront about the lack of evidence about exactly how extensive, reliable and frequent the access is — 'they are', as he says, 'not guinea pigs, after all'(Langland-Hassan 2015, p. 7)). A few extra details are still required to make the twins count as a threat to *introspective immunity*. A violation of that claim would require one twin to incorrectly self-ascribe a mental state otherwise accurately apprehended through introspection. It would require, that is to say, that the correctly discerned introspected mental state in fact occurs outside the introspecting twin's own mind, and yet that she is disposed to self-ascribe it on the basis of introspection. None of this is yet secured by the above minimal description of the twins' abilities, and it is unlikely to be settled by behavioural observation alone.

It was noted above that it is sometimes built into the very conception of introspection that it is a faculty by which we have access to our own mental states only. This might tempt us to the view that a twin's becoming introspectively aware of a mental state is enough to guarantee that the state is her own — a principle Langland-Hassan labels *Midas Touch*.[6] Following this suggestion to

[6]See (Langland-Hassan 2015, p. 12). His formulation is: 'Subject *S*'s becoming introspec-

its end would provide us with a reason for ruling out a reading of the case on which the introspecting twin really does introspect a mental state that is not her own — the very act of introspecting, on this view, would be enough to make it hers. The suggestion that we introduce an mental state ownership condition into the very definition of introspection, however, would get this result by stipulation alone; it is only because introspection has been designated as a faculty by which we have access to our own mental lives and none other that the introspecting twin cannot be counted as accessing a mental state that is not her own. The interest of such a result will quickly be drained of any substantive interest. Do we have a deeper reason to rule out such a reading?[7]

I think that we do once we bring in phenomenological considerations about the introspective experience. If their case is to be a genuine counterexample to *introspective immunity* then Krista's introspective access to Tatiana's first order mental states must be phenomenologically indiscriminable for her from her introspective access to mental states taken uncontroversially to be her own. One of the requirements, after all, was that Krista would be liable to (erroneously) self-ascribe the state on the basis of the introspective episode alone. For this to be the case, it must seem to her just as if she was engaged in a normal act of introspecting her own mental states. Langland-Hassan agrees. He registers the assumption that

> what it is like for Krista to be introspectively aware of [...] Tatiana's visual experience as of a toy pony [...] is more or less the same as what it is like for Tatiana to be introspectively aware of the same token experience. [...] [I]t is not a sort of "blind sight" counterpart to introspection — where she

tively aware of *m* suffices for *m*'s occurrence within *S*'s mind.' There is a potential ambiguity to this formulation: it is left unclear whether the sufficiency condition is intended as a bare modal claim (i.e. that whenever *S* is in a position to introspect *m*, *m* will be in *S*'s mind) or something like a stronger metaphysical constitution claim (i.e. it is *in virtue* of *S*'s being in a position to introspect *m* that *m* counts as occurring in *S*'s mind). That the formulation makes only a sufficiency, and not a necessity claim suggests the first reading, but the label *'Midas Touch'* suggests that he might have had something like the second reading in mind. I take it that the former is the more plausible contender, so — as I hope will become clear — that is the reading that I have in mind, and ultimately argue for, in this section.

[7] Langland-Hassan considers and rejects a number of underlying, non-stipulative reasons to go in for *Midas Touch*. Other than the response I favour, there is no room to go into a discussion of these suggestions here, but see §5 of his paper.

> simply finds herself with opinions about Tatiana's experiences, for reasons she knows not (Langland-Hassan 2015, p. 17)

We must assume, then, that any introspective phenomenology is just the same for Krista whether or not what she is engaged in is an ordinary episode of introspection. Given that the Hogan twins in fact seem perfectly capable of tracking ownership of mental states, it seems unlikely that their case is really one in which this phenomenological indistinguishability requirement is met. If, however, it *is* right that Krista's introspective experience really is subjectively indistinguishable from Tatiana's — or, more broadly put, from an episode of ordinary self-directed introspection — then we seem to have a reason to uphold *Midas Touch*. That is because if Krista's introspective awareness of a state is just like introspective awareness of an (undisputed) state of her own, then there is nothing to block it from becoming instantly absorbed into her mental life. This is because in the absence of a phenomenologically noticeable marker of non-ownership, she will have no way to classify an introspected state as her own or her sister's on the basis of the introspective experience alone. Given this lack of a discriminatory capacity, along with the related assumption (required for the counterexample) that the introspecting twin will be disposed to self-ascribe states on the basis of introspection, it is quite natural to think that she will, by default, be primed to incorporate any introspectively apprehended states into the rich and complex psychological fabric of her own mental life — into her standing beliefs, her immediate affordances for action, her long-term and short-term motivational profiles, and so on. It is hard to see, if this is right, why we *shouldn't* then treat it as a mental state of her own.

Things might not always, of course, go so smoothly. There might be some introspected states that will seem so discontinuous, for the introspecting twin, from the rest of her psychological life that she might be tempted to draw other-ascribing conclusions about its origin: an unannounced feeling of violent frustration intruding on a relaxed state of mind, for instance, or a highly detailed judgment that pops up, fully-formed, out of the blue. There are three things to say about such cases. The first is that it does nothing to throw out the claim that the state is now one of the introspecting subject's own; it is merely a state

that (given background knowledge of the situation) is recognised as having an unusual causal history, and from which the twin will presumably wish to dissociate herself. The second is that we should not be overly impressed with the grounds with which the misfittedness of these states would provide for the conclusion that the state in question began its life in her sister's mind. Even for non-conjoined conscious subjects, such mental jolts and interruptions are not at all unfamiliar.

To see these first two points, compare the following scenario. I have, let us suppose, been haunted all morning by a shadowy but unshakable feeling of sadness that I cannot pin down or easily account for. Eventually I pull myself together and confront my state of mind directly, making a real effort to trace it back to its probable trigger. With a little concentration, I finally come to the realisation that I am feeling this way because of the weepy film I watched last night. Several things might now happen. Perhaps the realisation will be enough by itself to dissolve the feeling of sadness. If it does not, I might choose to distance myself from the persisting state; after all, it's not as if I'm *really* sad, as I might put it. It's not a sadness with the same kind of causal history as my normal feelings of sadness. I might also, if it continues, begin to develop some second order attitudes about my state of mind — a certain kind of enjoyment in my own sadness, a self-directed schadenfreude, perhaps, or a feeling of vague amusement at seeing the effect that the film has had on me. Even if there is a sense in which I do not allow this state to become fully integrated with my current psychological profile, however, there is one thing that I would not do. I wouldn't refuse to recognise it as a state of my own, even if it is one that came about in a slightly unusual way. By themselves, then, cases like this give us no reason to yet treat Krista and Tatiana's case any differently.

But thirdly, and most importantly, the results of this sort of monitoring process — the process by which one twin recognises and rejects an introspected mental state as being out of character with the rest of her mental life and activities — could not restore the force of the case against *introspective immunity*. This is so even if it were somehow shown (contrary the first two points) that it did supply evidence that the introspected mental state was not the subject's own. That is because the twins' status as a counterexample depended on the

error of misidentification being made *on the basis of the introspective experience alone*. To introduce inferential considerations coming from such a monitoring process would be to shift the epistemic grounds.

In his paper, Langland-Hassan briefly considers a similar response to the one just presented. More specifically, he considers the following claim:

> *Strong Integration*: if mental state *m* is strongly causally and informationally integrated with subject *S*'s other mental states, then *m* occurs within *S*'s mind. (p.17)

Attraction to this claim, he suggests, might either come from a neo-Lockean theory of personal identity, or from the extended mind hypothesis. Both, he argues, are bad reasons to accept it. On the neo-Lockean view of personal identity he has in mind, *Strong Integration* would result in the unpalatable verdict that mere changes in perceptual state could occasion changes in personal identity. The extended mind hypothesis, by contrast, could only provide support for *Strong Integration* if we build into the case a reliability assumption. But, as he says, '[w]e can simply note that it is a logical (and, for all we know, nomological) possibility that the thalamic link provides for a kind of intermittent, unreliable access by each twin to the other's visual experiences.' (p.19)

These considerations against neo-Lockeanism and the extended mind hypothesis as incentives for accepting *Strong Integration* might well be right. But they also seem to be beside the point. Something like *Stong Integration* is, I think, compelling in its own right — an option that Langland-Hassan does not consider. Once a state has the same psychological currency as a subject's other mental states, that seems like reason enough to treat it as one of her own alongside the others. And as long as we maintain that there is no phenomenological difference between introspecting this state, and introspecting any other, then there is little reason to insist that it would not enter into the introspecting subject's cognitive economy. The principle at the heart of this way of responding to this case relies on the idea that the characteristic functional role of a mental state suffices to establish ownership facts about it: for the functional role of a state to include spontaneous rational transactions with a subject's other mental states, events and processes is enough to establish that the state is her own.

It might not be bulletproof, but at the very least it would seem ungenerous to deny that this principle is compelling on its own terms; we do not need to treat this approach as the product of neighbouring theoretical commitments to see its attractions.

Cases of craniopagus twins, then, do not undermine the immunity to error through misidentification of judgments formed on the basis of introspection. I have argued that so long as the introspecting twin is in a position to self-ascribe the state, that state will be playing a role in her cognitive economy that legitimises the self-ascription, and so the judgment will not be in error through misidentification. As long as we accept this picture of mental ownership, we have a non-stipulative reason to hold that the very act of introspecting a state is enough to ensure that it is a state of one's own.

4.2 Memory

Episodic memory is the faculty by which we normally come to know about how things were with us in the past.[8] Through memory, information encoded at an earlier time is rendered available for retrieval at a later point. It stands out, in this respect, from the other forms of self-knowledge considered in this chapter, in that memory is not generative of self-knowledge, but preservative of self-knowledge gained in other ways and at other times.

Memory is a multiply fallible faculty even in its best light, and in its worst, 'a notorious deceiver' (Martin 1992, p.752). The question that concerns us now is whether one of the many faces of that fallibility is a vulnerability to errors of misidentification relative to uses of the first person concept. To answer in the negative is to sign up to the following principle:

> *Memory immunity:* I cannot be wrong, in forming a judgment about a past state, event or process of the form *I was F (F-ed, was F-ing)* on the basis of episodic memory solely in virtue of a misidentificatory mistake about whether or not it is *me* that I know to have been *F* on those grounds.

[8] There are many different kinds of memory; my interest in this section is exclusively with autobiographical episodic memory.

Probably no one opposes *memory immunity* under a *de facto* reading. In our world and for creatures like us, we couldn't make errors of this kind. This means that any challenge to *memory immunity* must be one of a merely conceivable kind. The principal conceptual challenge comes in the form of so-called quasi- or Q-memory, first introduced by Shoemaker. This is the describable possibility of memory impressions that are derived from somebody's past experiences in much the same way in which our ordinary memories are derived from our own past experiences, with the only difference being that the person from whose past experiences the impressions derive might or might not be identical with the remembering subject. If this is a coherent possibility, it would be, as Shoemaker writes,

> a kind of knowledge of past events such that someone's having this sort of knowledge of an event does involve there being a correspondence between his present cognitive state and a past cognitive and sensory state that was of the event, but such that this correspondence, although otherwise just like that which exists in memory, does not necessarily involve that past state's having been a state of the very same person who subsequently has the knowledge. (Shoemaker 1970, p.172)

The conceivability of quasi-memory, the idea is, demonstrates that it is at least in principle possible to use memory as a source of self-ascriptive judgments that are in error through precisely the kind of mistake ruled out by *memory immunity*, even if there could be no such judgments as things really are for us. That is because ordinary memories are just a special kind of quasi-memory, and quasi-memories are a way that we have of knowing about how things were in the past with *someone*, if not with ourselves. The aim of this section is to defend the immunity to error through misidentification of first personal memory-based judgments against this challenge coming from the conceivability of quasi-memory.

First, though, a quick methodological note. The combinatorial nature of the knowledge gained through memory raises special difficulties relating to the isolation of the epistemic profile of remembering. Given that the final judgment is a product of the interaction between memory and other initial knowledge sources, a final verdict of vulnerability to error through misidentification cannot by itself deliver the point at which the potential for error was introduced.

As Shoemaker observes, the way to avoid this difficulty is to hold fixed the immunity to error through misidentification of the earlier judgment, *I am F*, based on the original source of self-knowledge; the immunity or vulnerability to misidentification error of the later memory-based judgment, *I was F*, will then be due to the faculty of memory alone.[9]

4.2.1 Evans vs. Shoemaker

The principal cases for and against the efficacy of the quasi-memory challenge against the immunity to error through misidentification of memory-based first person judgments can be helpfully set out using the respective positions of Shoemaker and Evans. The case 'for' is given by Shoemaker, who simply points to the conceivability of quasi-memory as defined above. The significance of this conceivability, Shoemaker explains, is that our knowledge of our pasts is shown to be less direct than we might have thought; the conceivability of quasi-memory shows our memory-based judgments to be grounded in a criterion of identification, even if that is something that normally goes unnoticed as things are for us. He writes:

> [S]uppose that it were possible to quasi-remember experiences other than one's own. If this were so one might remember a past experience but not know whether one was remembering it or only quasi-remembering it. Here, it seems, it would be perfectly appropriate to employ a criterion of identity to determine whether the quasi-remembered experience was one's own. [...] Thus the question of whether the knowledge of our own identities provided us by memory is essentially non-critical turns on the question of whether it is possible to quasi-remember past actions and experiences without remembering them. (Shoemaker 1970, p.272)

The possibility of quasi-memory shows memory-based judgment to involve an identification, even if such judgments have *de facto* immunity to error through

[9]See (Shoemaker 1970, p.270); Bermudez has recently made a similar point, (Bermudez 2013). Evans handles this point slightly differently by suggesting a refinement of the notion of immunity to error through misidentification: 'We need, therefore, a more sophisticated notion of immunity to error through misidentification than before: one which can highlight the absence of an identification of a certain kind, or at a certain point in the process that issues in a judgement, rather than the absence of any identification at all' (Evans 1982, p. 238).

misidentification in our world.

Granting the conceivability of quasi-memory, Evans argues by contrast, does not force us to concede the vulnerability to error through misidentification of ordinary memory-based judgments. That is because such a concession would force us to say that the deliverances of memory are person-neutral. It would force us, that is to say, to the view that if it does not provide a subject with evidence about how things were with *her* in the past, a memory could nevertheless provide her with evidence about how things were with *someone*; 'The picture requires us to be able to think of memory as a way of having knowledge of an object which leaves its identity [...] an open question.'

Evans dispenses with this suggestion with only a very quick comment, that 'it does not appear to me to be possible to think of memory as an 'identity-neutral' way of having knowledge of the past states of a person' (Evans 1982, p244). The objection is one of implausibility; Evans takes it to be doubtful that the deliverances of memory could ever be neutral on the question of whose past it is from which they derive. Memory only ever provides me with (mis)information about how things were in *my* past, even if causal deviant chains like the ones Shoemaker describes might lead to systematic illusions about how things were in that past. Even in a case of quasi-memory the content of the memory impression involves a representation of *me* as the subject of the past event. Such cases, then, rightly demonstrate the possibility of memory-based error, but not an error of misidentification — I am not making a mistake about who the remembered events happened to, only about what things happened to me.

In 'Error through misidentification: some varieties' (2006), Annalisa Coliva convincingly draws out the equal appeal of these two positions. On the one hand, it seems right that if we accept the coherence of quasi-memory, then we must allow the possibility of forming a judgment on the basis of a memory impression in which I am right about *what* happened, only wrong about *who* it happened to. But on the other hand, even in such a case, it seems right that the judgment is not formed via an identification; the memory impression gives me immediate warrant to form a judgment about how things were with *me* in the past, it's just that under these conditions I will be systematically wrong.

As she says, '[i]t would perhaps be an overstatement to call this opposition a paradox, but it is undeniable that both theorists seem to have a point.' (Coliva 2006, p.407) Coliva goes on to refine the notion of immunity to error through misidentification in an attempt to resolve the tension between them, a strategy inherited from Pryor's earlier paper 'Immunity to error through misidentification' (1999). An alternative approach, taken up by authors such as Jordi Fernández, is to find a way of breaking the tie.[10] I will do this by drawing on recent work in the cognitive sciences in support of Evans's position.

4.2.2 Storehouse vs. constructive paleontology

A dominant model of memory in contemporary philosophical discussions can be retraced to the seminal paper, 'Remembering' (1966), by C. B. Martin and Max Deutscher. With a number of qualifications pertaining to the third clause, they offer three conditions that must hold of a remembering subject:

1. Within certain limits of accuracy he represents that thing.

2. If the thing was "public", then he observed what he now represents. If the thing was "private", then it was his.

3. His past experience of the thing was operative in producing a state or successive states in him finally operative in producing his representation. (Martin and Deutscher 1966, p.166)

The qualifications are introduced to rule out deviant ways in which the past experience might be operative in producing the current representation, but will not matter for our purposes. What *is* key to understanding Martin and Deutscher's third clause is the idea of a *memory trace*, a structural analogue of what was experienced that is laid down at the time of experiencing, retained over time, and reactivated at the later time of remembering. To illustrate the point, Martin and Deutscher appeal to the familiar image of a storehouse; they write, '[s]o long as we hold some sort of "storage" or "trace" account of memory, it follows that we can remember only what we had experienced, for it

[10](Fernández 2014)

is in our experience of events that they "enter" the storehouse.' (Martin and Deutscher 1966, p.189)

It is, I think, no exaggeration to see this storehouse model as thoroughly antithetical to the model of memory on which the cognitive sciences have converged over the last century or so. This is the reconstructive model of memory, that has emerged — following pioneering experimental work by the psychologist Frederick Bartlett in the 1930s — as a 'modern consensus' (Sutton 2011, p.355) in psychology, the neurosciences and clinical practice. As psychologists H.L. Roedinger and K.A. DeSoto write, '[c]ontrary to popular belief, memory does not work like a video-recorder, faithfully capturing the past to be played back unerringly at a later time. Rather, even when our memories are accurate, we have reconstructed events from the past' (Roedinger and DeSoto 2015, p.1). Rather than a simple reactivation of dormant causal traces, psychologists tells us, memory is an active reconstructive process, heavily influenced by a daunting range of factors. These include 'hormonal and neuromodulatory, genetic and pharmacological, developmental and age-related influences; there are influences of arousal, stress, gender, mood, emotion, sleep and personality; there are unconscious, schematic or semantic influences, and there are influences of context, situation, task, and environment.' (Sutton 2011, p.355)

This reconstructive model has a metaphor of its own:

> To draw on a metaphor from Hebb (1949), we can think of the process of remembering the past as we conceive of the paleontologists' reconstruction of a dinosaur from bone fragments and chips. The archeologist recovers a partial skeleton, but the finished product in a museum is shown as complete, with new bones added, old ones refinished or enhanced, and the entire skeleton reconstructed based on knowledge of what the animal probably looked like. (Roedinger and DeSoto 2015, p.10)

Of course, this paleontological model does not deny the involvement of causal traces in the activity of remembering — these are the 'bone fragments and chips' in the picture above. In the good case at least, there is plainly a causal connection between the past event and the present memory, and if there is to be no causation at a temporal distance then this must be due to encoded traces

of some kind. But these traces are given a relatively minor role in the production of the final memory. As Sutton and Carl Windhorst explain, 'even when there are causal connections between events and traces, and between traces and remembering, these connections are multiple, indirect and context-dependent. [...] Rather, we think that for personal memory at least it is plausibly the norm that some details crop up in remembering an experience which have not been encoded in the same trace as that experience: indeed, that the idea of one trace per experience is both conceptually and empirically highly dubious.' (Sutton and Windhorst 2009) In addition to material from the original memory trace, the final memory is likely to include material from other memory traces, along with additional material from the imagination and the operation of a process called 'pattern completion' — roughly, filling in the details — that is enabled by background knowledge and expectations about the kinds of events remembered.

There is little remaining doubt as to which of these models is our best current empirical theory. (Sutton evocatively describes the sense of being 'caught up in the righteous rejection of misguided archival models of memory as extraction of untainted materials from an inner storehouse'. (Sutton 2011, p.356)) I now want to suggest that something like the storehouse model sits in the background of Shoemaker's treatment of quasi-memory, and that taking the well-advised move to the reconstructive model should bring with it a shift towards Evans's approach.

4.2.3 Using the reconstructive model of memory against the case for quasi-memory

The question we now face is, in a case of remembering in which the dominant memory trace causally derives from the past experiences of someone other than the remembering subject, would this show the memory-judgment to be vulnerable to error through misidentification? In self-ascribing the past experience, would the remembering subject be making a mistake solely in virtue of an error about whether or not it's *she, herself* to which she knows the past-tensed predicate to apply on the basis of memory?

Let's quickly remind ourselves of the two positions set out above about quasi-memory. Briefly put, Shoemaker's view was that the conceivability of quasi-memory demonstrates the possibility of making a memory-based judgment in which a subject does know, on the basis of memory, that *someone* ϕ-ed, but is wrong about whether it was her; the conceivability of quasi-memory straightforwardly reveals our memory-based judgments to be vulnerable to error through misidentification relative to uses of the first person concept. Evans disagreed. The conceivability of quasi-memory does not undermine the immunity to error through misidentification of first personal memory-based judgments, Evans argued, because memory is not a faculty whose deliverances could ever be person-neutral. Memory is only ever a source of information, or misinformation, about how things were with *me* — the conceivability of quasi-memory shows only the possibility of conditions under which these deliverances would be systematically illusory.[11]

Which is the right response to our question? On the storehouse model of memory, it is easy to be seduced by Shoemaker's answer. What we are being asked to imagine, on this model, is an earlier experiencing subject laying down a memory trace that is then artificially implanted into a second subject's brain to be reactivated at a later occasion. There is nothing in this construal of the situation that obviously prohibits a description of the remembering subject's later memory-based judgment as one that is in error solely through an error of misidentification. Indeed, if the remembering subject was to form the judgment, on the basis of a causally deviant memory impression, that *I was F*, this — at least prima facie — seems to be precisely a case in which the subject has knowledge that the predicate applies, but is making a mistake in judging that it applies to her.

Even accepting this traditional storehouse model of memory there are ways of resisting Shoemaker's verdict,[12] but I want to try something else. I want to see what happens when we replace it with a reconstructive model of memory. To begin with, it seems that if we take memories to be constructions made out

[11] Evans allows that this systematic illusion could itself form the basis of a source of knowledge about the other subject's past, but only as a form of *inferential* knowledge, rather than pure memory-based knowledge.

[12] See, e.g. (Roache 2006)

of materials provided by multiple memory traces, imagination, background knowledge, expectations and an abundance of contextual and physiological influences, then we are being asked to imagine something rather different in the case of quasi-memory. The story, again, begins with an earlier experiencing subject laying down a memory trace that is artificially implanted into a second subject's brain. Things look very different, however, when we get to the point of the remembering. The remembering subject does not simply and cleanly reactivate the causally deviant memory trace. Rather, she engages in a creative process of construction that draws on that trace, but that also draws on materials of many other kinds, including detailed (though mostly unconscious) background knowledge of her own history, personality, goals, beliefs, interests and expectations. These are what Bartlett originally called *schemas*, the 'active developing patterns', or 'active organisation of past reactions, or past experiences' that scaffold the reconstructive process of remembering. (Bartlett 1932, p.11)

With the case so construed, it is far less tempting to view the mistake the remembering subject is making as one of pure misidentification. The reconstructive process is profoundly fallible in many interesting and surprising ways that have sometimes been argued to serve adaptive purposes, or at least to be the side products of other adaptive features.[13] Indeed, one of these ways in which memory reliably fails us is through errors of misattribution — quasi-memory cases of a kind, in which we mistakenly attribute a recollection to the wrong source.[14] Among these diverse potential errors, the causally deviant (but minimally contributive) memory trace built into the quasi-memory case starts to look rather less impressive. That is because no matter what mistakes are involved in the constructive process of memory, one thing seems to be held fixed: the materials and processes employed all seem to be ineliminably self-specific, in that they all contribute towards putting together a picture of the remembering subject's own past. That is because no matter how distortive or gappy the process is, the biological function of remembering is one of providing information (or misinformation) about how things were *with the remembering subject* in

[13](Schachter 1999)
[14]See, e.g. (Schachter 1999, pp.188-191)

the past.

Indeed, although he does not talk about biological function, this is very close to a point made by Bartlett himself:

> Thus what we remember, belonging more particularly to some special active pattern, is always normally checked by the reconstructed or the striking material of other active settings. It is, accordingly, apt to take on a peculiarity of some kind which, in any given case, expresses the temperament, or the character, of the person who effects the recall. This may be why, in almost all psychological descriptions of memory processes, memory is said to have a characteristically *personal* flavour. If this view is correct, however, *memory is personal, not because of some intangible and hypothetical persisting 'self' which receives and maintains innumerable traces, re-stimulating them whenever it needs; but because the mechanism of adult memory demands an organisation of 'schemata'* depending upon an interplay of appetites, interests and ideals peculiar to any given subject. (Bartlett 1932, pp.9-10, emphasis added)

Memory is not person-neutral. This is not because it is inconceivable that a contributive memory trace could have causally derived from elsewhere, but because the reconstructive process of memory is filtered through self-specific schemata; in forming a memory, I draw on an enormous repository of variously sourced (mis)information *about myself* to form a picture about how things once were with me.

In the light of this prevailing model of memory in the cognitive sciences, then, we should agree with Evans that the conceivability of quasi-memory is not enough to overturn *memory immunity*. The conceivability of these cases only serves to demonstrate that in addition to many natural pitfalls of memory, there could also be non-natural ways of inducing systematic illusions about how things used to be with the remembering subject.[15]

[15]One writer who argues for something very close to this idea is Marya Shechtman. Her work on memory focusses on the idea that the content of a single autobiographical memory partly depends on its place in a broader psychological context, and that it is only as part of that broader context that the memory will be shot through with autobiographical or first personal significance. See, e.g. (Schechtman 1996) and (Schechtman 2010).

Chapter 5

Which forms of bodily awareness?

The aim of the last chapter was to defend the immunity to error through misidentification relative to uses of the first person concept of judgments based on two forms of self-knowledge we might naturally classify as mental; introspection and memory. The aim of this chapter is to do likewise for two forms of self-knowledge usually classified as bodily. These are the related faculties of so-called internal bodily awareness and multimodal bodily awareness. There is a faint flavour of artificiality to this separation. All four of these faculties are available to us as forms of self-knowledge only because we are conscious creatures who experience ourselves in each of these various ways. There is a sense, then, in which all of them are mental forms of self-knowledge. On the other hand, there is also a sense in which all four faculties are only available to us because we are embodied creatures, at least on a non-Cartesian model of the mind. A different way of categorising them, then, would cast them all as bodily forms of self-knowledge. I have organised the cases this way over these two chapters merely for ease of presentation.

5.1 Internal bodily awareness

We have many ways of perceiving our bodies 'from the inside' which we might naturally suppose are exclusively available for perceiving our *own* bodies only. Gareth Evans lists these as 'our proprioceptive sense, our sense of balance, of heat and cold, and of pressure' (Evans 1982, p.220), though we might want to

add to these, among others, our kinaesthetic sense, haptic sensations, nociception, and sensitivity to itches, tickles and their like. This collection of internal sense modalities is our faculty of *internal bodily awareness*.

If it is right that internal bodily awareness is an epistemic source that is dedicated to a single object — that is, to one's own body — then judgements made on it basis will be immune to error through misidentification. I might be mistaken about what it is that I am so perceiving (say, I might mistake an intense itch for a pain), I cannot be mistaken solely in virtue of being wrong about whose body it is that I am perceiving. The single-object dedication of my faculty of internal bodily awareness ensures that those informational deliverances could not fail to derive from my body alone.

Such considerations bring us to this:

> *Internal bodily awareness immunity:* I cannot be wrong, in forming a judgment about an occurrent physical state, event or process for the form *I am F* on the basis of internal bodily awareness solely in virtue of a misidentificatory mistake about whether or not it is *me* that I know to be *F* on those grounds.

There are both conceptual and empirical challenges to *internal bodily awareness immunity.* I will consider each in turn.

5.1.1 Redirected wire cases

The case

One of the principal conceptual challenges to this thesis comes in the form of what we might call *redirected wire cases* — thought-experiments in which a subject is wired to another's body in such a way that she receives internal bodily awareness from the other body that is phenomenologically indistinguishable from an episode of normal internal bodily awareness.[1] Take the following example from Lucy O'Brien:

[1]For examples of this challenge see, e.g. (Armstrong 1984, p. 113), (Martin 1995, pp. 275-6), (Evans 1982, pp. 221, 250), (Cassam 1997, p. 63), (O'Brien 1995), (O'Brien 2007, p. 206), (O'Brien 2012)

> Imagine that sometime in the future baby products manufacturers provide us with the new 'Internal Baby Monitor' (the IBM). In the case of a screaming child, one sticks the device to the baby and to oneself and one is then presented with the baby's body space 'from the inside'. [...] Given regular use of such a device it is not hard to imagine circumstances when I wake — having gone to sleep with this gadget on — and wonder whether I have my leg bent over, or the baby has, on account of not being sure whether the device was on or off. (O'Brien 2007, p. 206)

Based on the experience just described the perceiving subject in this case might make a judgment on the basis of internal bodily awareness. Unlike our ordinary exercises of that faculty, however, it looks like this subject is in danger of misidentifying the object of her judgement. She might be in error, that is to say, in the very way that would violate *internal bodily awareness immunity* — by going wrong solely in virtue of misidentifying the body she is perceiving.

There are at least two ways of understanding the threat posed by these cases. The first is the weakest. On it, redirected wire cases serve to show that although internal bodily awareness is the source of epistemically secure identification-free judgments as things are for us, this is just a contingent aspect of our empirical circumstances; things need not have been so. This rendering of the threat leaves the identification-freedom of (and so the immunity to error through misidentification of the judgments based on) our faculty of internal bodily awareness intact. Redirected wire cases, on this reading, merely illustrate the observation that the world might have been otherwise. This version of the threat is really no threat at all.

The stronger version of the threat recognises the implications of the perceiving subject's predicament in the redirected wire case for our understanding of the structure of internal bodily awareness as it is for us as well as for her. So long as we insist that the faculty in play remains constant between her world and our own — an assumption on which the relevance of the case importantly rests — the metaphysical possibility of a misidentification error lays bare the identification-involving nature of that faculty. If, that is to say, internal bodily awareness is a modality that makes counterfactual room for the possibility for an identification error then it must, *a fortoriori*, be a modality whose internal

structure involves an identification, even if for creatures like us there is no *de facto* chance of it going wrong. Redirected wire cases, on this understanding of the threat, make salient the presence of an identification in internal bodily awareness where it had previously been hidden by the contingent privacy of the faculty as it is in our actual world. By these lights, internal bodily awareness is an identification-involving source of self-consciousness, albeit an especially reliable one in our world. As such it cannot be a form of self-knowledge that issues in judgments with immunity to error through misidentification relative to uses of the first person concept.

My response to this challenge to *internal bodily awareness immunity* comes in two parts. In the next section I consider a number of ways in which the threat from redirected wire cases might be blocked. The arguments of that section are not conclusive. But they serve to demonstrate a crucial point: the coherence of efficacious redirected wiring scenarios cannot be simply assumed, it must be earned through careful composition of the cases' details. The way of finally deflecting the challenge comes in the last subsection of §5.1.1. I argue that no way of filling in the details will serve to uphold the threat to *internal bodily awareness immunity.*

Filling in the details

What would it be like to perceive another body's pain, tickle or limb-bend 'from the inside'? To answer this question, authors of redirected wire cases face the following choice: any internally perceived bodily state or event experienced as having a location must be experienced as falling either inside or outside the subject's felt bodily boundaries. This, of course, might not have been so had we not been the sorts of creatures that are conscious of the limits of our bodies. Given, however, that we *are* such creatures, those felt boundaries must be accounted for in fully specified descriptions of redirected wire cases. Do those felt boundaries encompass the apparent location of the object perceived through the redirected wiring system or not?[2]

[2]It is important to note that these are only the *experienced* boundaries of one's body, experience which need not be veridical —- an amputee's felt bodily boundaries, for instance, might encompass an absent 'ghost' limb.

One possible line of response is to press for the presence of a dilemma in this choice. Neither answer, it might be argued, could uphold the intended force of these cases.

How would the dilemma go? To see the first horn, suppose that the case is described in such a way that the subject feels the limb-bend or tingle to be located *outside* her felt bodily boundaries. It is a condition on the success of redirected wire cases that the redirected wire experience of the other body is phenomenologically indistinguishable for the subject from an ordinary episode of self-directed internal bodily awareness. (We have seen an analogous version of this phenomenological indistinguishability requirement in the case of introspection in §4.1.2.) After all, if the redirected wire experience differed noticeably from an experience of ordinary internal bodily awareness then the subject would have available an easy method of discrimination between the two kinds of experience. She would then, however, have no inclination to form a first personal judgment on the basis of redirected wire input — that it was not her body in question would be phenomenologically salient to her. Describing the perceiving subject of the redirected wire set-up as experiencing the object of her redirected wire perception to be located outside her felt bodily boundaries, then, would be to place the case in conspicuous violation of this phenomenological indistinguishability condition. The perceiving subject could now simply read off from the felt location of the sensation whether it is the other body or her own that she is perceiving. She would have no call to respond to redirected wire sensory input with a first personal judgment, and so no reason to form a judgment in error through misidentification. This is the first horn of the proposed dilemma for the redirected wire proponent.

The second option is to describe the internally perceived state or event as being experienced as falling *within* the subject's felt bodily boundaries. This description of the case has a clear advantage — at least with respect to the question of containment within her bodily boundaries we can be confident that the redirected wire experience will be indistinguishable for the subject from an episode of ordinary internal bodily awareness. There is no reason to doubt that the condition of phenomenological indistinguishability has been met.

The problem for this second option, however, is that this reading of the

case contributes no special resources for establishing the claim that the perceiving subject really is enjoying perceptual access to the other body. What is prescribed by the set-up of the case is only that the perceiving subject's experience is, or could be, *caused* by states and events in the second subject's body, but to say this much is not yet to determine that her experience is *of* that body. Indeed, given that she feels the sensation to be located within her own bodily boundaries, the more likely explanation is that a sensation in the second body has caused the perceiving subject to have a sensation in her own. The experiment's wiring system is, to borrow a line from M.G.F. Martin, 'a sophisticated mechanism for causing [sensations] in two people instead of one' (Martin 1995, p. 276). And on this reading of the situation there would, of course, be no mystery about the redirected wire experience's phenomenological likeness to an ordinary episode of internal bodily awareness — the perceiving subject really *is* engaged in a state of ordinary self-directed internal awareness (albeit one with an unusual causal history). Under this description of the case the subject's first personal judgment will not be in error through misidentification; the judgment really is about herself.

It might be objected that this move overlooks the fact that there are different ways in which we might envisage the operation of the wiring system linking the two subjects. On one construal, the redirected wires function by simulating nerve ending stimulation in the perceiving subject's relevant body-part. On this understanding of the set-up it is difficult to see how the suggestion of the last paragraph — that the sensation in the second body causes an independent sensation in the first — could be resisted, but an alternative rendering of the case might not be so easily dismissed. Perhaps we are to understand the wires as operating by artificial stimulation of the relevant areas of the perceiving subject's cerebellum. Are we similarly forced, on this understanding of the mechanism, to the view that she is undergoing an ordinary internal perception of her own body? Perhaps not. But such a move should be of little encouragement to supporters of the case; the sensation in the second body is now a mere causal antecedent of a neurologically induced illusion of a sensation in the perceiving subject.

Either way, then, the case fails to secure the result that the subject's first

personal judgment is, or could be, in error through a mistaken identification. On the first construal of the set-up, the bodily event in the second body merely causes another event in the first, so the judgment will not be in error through misidentification. On the second, on the other hand, the judgment is based on an illusion — an erroneous judgment for sure, but not in virtue of a misidentificatory error. This, then, is the second horn of the proposed dilemma. In opting to describe the sensation as falling within the perceiving subject's felt bodily boundaries, the description of the case fails to establish the possibility that the subject's judgment is in error through misidentification.

Again, the redirected wire proponent might object to the last suggestion, that the second construal of the set-up would amount to a neurologically induced illusion of a sensation in her own body. Why *shouldn't* an appropriately caused stimulation of the experiencing subject's brain count as a genuine perception? What we have here, after all, is a sensory causal link that provides the subject with reliable information about how things are with the second body. If such considerations serve to at least make room for an intuition that the experiencing subject really does perceive the second body then surely the cases' supporters have the right to insist that it is the task of the cases' objectors to rule this possibility out, rather than that of their proponents to rule it in.

This is just the task taken up by Martin in 'Bodily Awareness: a sense of ownership' (1995), in which he argues that the sense of ownership over our bodies that we experience in internal bodily awareness gives us reason to think that perceptual access through that faculty is restricted to our own bodies only. No matter how reliable the causal link, Martin argues, redirected wire experiences cannot be genuinely perceptual.

On Martin's view, the sense of ownership is a phenomenological feature necessarily accompanying all episodes of internal bodily awareness, or at least all episodes whose perceptual objects are experienced as having a location. This is because to feel a bodily state or event as located, on Martin's view, is to feel it as located in a region of space within which one *could* perceive bodily sensations, and this is just what it means to feel it to be located within the bounds of one's own body. As such, 'the quality of falling within one's bodily boundaries is not independent of the felt location of the sensation. The sense

one has of the location of sensation brings with it the sense that the location in question falls within one of one's apparent boundaries.' (p.271)

Martin observes that if the link involved in redirected wire cases was sufficient for perception then this sense of ownership could not be relied upon to be veridical. All (located) bodily states and events perceived through such a link would be characterised by a sense of ownership, but in these cases that sense would be misleading; they would, in fact, be perceptions of a body other than the perceiver's own. This is to say that if redirected wire perceptions were allowed, then the phenomenology of bodily ownership would fail to track the facts of bodily ownership — the sense of ownership could not be treated as a genuinely perceived property of sensations. This is because even in a good case, the fact that the sensation is in fact occurring within one's own body would be nothing more than accidentally related to the sense of ownership; 'there is no perceptual connection between the body-part seeming so and its actually being so' (p.278). It is the embarrassment of this result that leads Martin to his conclusion: to uphold the intuition that the sense of ownership is a genuinely perceived property of our perceptions through internal bodily awareness, it must be denied that a causal connection to another body could ever be genuinely perceptual.

There are a number of ways of challenging Martin's argument. One is to consider more carefully the argument's central notion of the sense of ownership.[3] O'Brien, for instance, has argued against the claim that such a sense of ownership is an essential feature of all sensations, located or otherwise. Given the way things are for us, she allows, it might seem as though internal bodily awareness is essentially imbued with a first personal sense of ownership. Such an impression, however, need be nothing more than a bias born of background beliefs about our empirical circumstances. Take again her example in which I am connected to an internal baby monitor, and in which I am unsure

[3]The other dominant source of resistance to Martin's argument comes from the question of how to reconcile Martin's identification of the sense of ownership with experienced locatedness with empirical cases of somatoparaphenia (discussed in §5.1.2); a delusion associated with neurological damage to posterior cerebral regions in which sufferers deny ownership over isolated body-parts despite displaying a capacity to accurately identify the location of sensations occurring in the affected area. See (de Vignemont 2013) for a full presentation of this objection.

whether it is my own leg or my baby's that is bent over, in virtue of being unsure whether the monitor is on or off. In a world of such doubts, the idea is, we can imagine a drift — perhaps over some time — in the person-relative content of internal bodily perception. Even if as a matter of fact I sometimes perceive my own body and sometimes my baby's, my beliefs about (and perhaps habituation into) the environmental context as being one in which redirected wire mechanisms are sometimes operative would eventually lead to both internal bodily awareness and redirected wire baby-body awareness gradually becoming person-neutral in content. The 'from the inside' nature of internal bodily awareness, O'Brien concludes, is in principle separable from its first personal or proprietary character. If this is right, then the question of whether the causal link involved in redirected wire cases suffices for perception cannot be settled under the terms proposed by Martin. The sense of ownership would be merely a contingent feature of bodily awareness as it is for us, a feature whose relevance would not extend to redirected wire worlds.

Neither Martin nor O'Brien deny the conceivability of a subject having internal bodily awareness of a body other than her own; their dispute hangs only on the question of whether such an experience would be necessarily accompanied by a sense of ownership. For Martin, it is because such experiences *would* necessarily be so accompanied that redirected wire experiences — while conceivable — cannot be genuinely perceptual. For O'Brien, on the other hand, it is the conceivability of redirected wires cases that establishes the eliminability of the sense of ownership characteristically accompanying exercises of internal bodily awareness in our own non-IBM world. I think that both sides of this dispute are too quick in allowing the conceivability of genuine internal bodily awareness of a body other than the perceiver's own — the question whether internal bodily awareness is necessarily attended by a sense of ownership, rather, must itself await the verdict on the viability of these cases. In the next section I will argue that there is no way of filling in the cases' details that could vindicate any such claim to viability.

The response

There is an important respect in which the faculty of internal bodily awareness is unlike the modality of vision. When I look down and see my ankle beneath me, I see it as located at a certain distance and angle along an axis originating from my eyes. This clearly differs from a nociceptive experience of the same body part. When I feel a pain in my ankle, there is no centralised physiological mechanism analogous to the optic system from which the pain is perceived. Nociceptive receptors, rather, are distributed more or less evenly throughout the body so that pain can be detected wherever damage occurs. Unlike vision, internal bodily awareness lacks a perspectival structure.[4]

This structural difference between vision and internal bodily awareness is hardly surprising given the anatomical differences between those sensory systems. In the case of vision, several million photoreceptors are gathered on the retina to absorb incoming light particles from outside the eye. By contrast, somatosensation — the sensory system encompassing the cutaneous senses for the detection of touch, pressure and vibrations, and the nociceptive, thermoceptive, proprioceptive and kinaesthetic systems — is comprised of receptors distributed with differing concentrations throughout the body, over muscles, bones, skin, epithelial and connective tissues and internal organs, which harvest information from any of those body parts. Our sensitivity to muscle stretch, for instance, is due to muscle spindles, or small encapsulated sensory receptors, positioned throughout each of our muscles in varying levels of density, which serve to detect any changes in the state of the muscle's extension. In such systems, the body and its parts serve at once as both the organ and the object of perception. No wonder, then, that their exercise lacks a perspectival structure.

A first step in defence of *internal bodily awareness immunity* against the threat from redirected wire cases is to notice that internal bodily awareness could not share its non-perspectival spatial structure with redirected wire experiences. This follows from the minimal commitments of the cases as described; redirected wire perceptions *must* be perspectival, given that the perceiving subject

[4]That internal bodily awareness has a non-perspectival structure is, of course, nothing new; see (Soteriou 2013), (Bermudez 2005), and (Smith 2002) for some recent discussions.

is *ex hypothesi* using that faculty as a way to remotely perceive a body other than her own — there must, this is to say, be a point of perceptual origin that is distinct from the location of the perceptual object; the origin is with the perceiving subject, the object with the perceived.

The very same point can also be put another way, which is to say that redirected wire experiences could not become non-perspectival in structure without a corresponding shift in perceiver. This is because only the second subject can actually perceive her own body 'from the inside' at no distance, so to insist that the sensation is experienced non-perspectivally would be to restore the second subject as the sensation's perceiver. Redirected wire perceptions of a body other than the perceiver's own, then, cannot bear the same spatial structure as genuine internal bodily awareness. This first step delivers a significant result: the apparent conceivability of redirected wire cases relies on the conceivability of a sense modality that is differently structured to internal bodily awareness.[5]

Earlier, in the first subsection of §5.1.1, we saw that the real threat of redirected wire cases relies on a crucial identity assumption about the epistemic grounds involved in the formation of ordinary internal bodily awareness judgments and those involved in judgments of redirected wire perception. That those same grounds sometimes make room for identification errors, the idea was, reveals them to be identification-involving, even if the identification is normally kept hidden by the contingent single-object dedication of the faculty of internal bodily awareness as it is for us. The problem now facing the redirected wire proponent is that if internal bodily awareness and redirected wire experiences differ in perspectival structure as we have just seen, then this identity assumption must be false.

There are two reasons for accepting this conditional. The first, quite straightforwardly, is that this perspectival-structural difference itself gives us reason enough to treat internal bodily awareness and redirected wire experience as distinct kinds of epistemic grounds. The identity assumption simply cannot

[5]Notice that there is no parallel such point to be made about redirected wire cases for the faculty of vision, because vision — even when all goes well — is already perspectivally structured. This means that in the case of vision there exists a distal relation between a point of origin and the objects perceived, a relation conceivably vulnerable to the kind of distortion needed for redirected wire cases.

survive such a radical structural difference. But there are ways of challenging this suggestion — is this difference really difference enough to undermine the identity assumption? After all, as Martin points out, 'we do not tend to think that the use of spectacles, or even the use of binoculars or telescopes, prevents us from genuinely seeing objects through them' (1995, p.277). Why think that a prosthetic device extending the range of internal bodily awareness beyond its normal scope will prevent a subject from enjoying a state of genuine internal bodily awareness, and so forming judgments on the very same kinds of grounds as ordinary internal bodily awareness judgments?

This objection can, I think, be met. The difference between internal bodily awareness and redirected wire experience just isn't like the difference between ordinary vision and looking through a telescope: it involves structural distortion rather than mere variation in scope. In general, it is difficult to know how to settle individuation questions for epistemic grounds, but in this case I think there is a clear reason for taking structural differences of this kind to be relevant to the interests of all parties to the present dispute. That is, that the redirected wire proponent is drawing on the identity assumption in the service of revealing the *identification-involvement* of those grounds. Identification-involvement, however, is a structural feature of judgmental grounds rather than, say, one of scope. We should therefore expect that a difference in a ground's structure will open the way for a difference in its identification-involvement even if a difference in scope will not.

Putting all this to one side, however, there is also a second reason to think that a difference in perspectival structure implies a difference in epistemic grounds. It is, we have already seen, key to the efficacy of redirected wire cases that the redirected wire perception of the second body is subjectively indistinguishable from an ordinary episode of internal bodily awareness. For this reason, any structural differences between internal bodily awareness and redirected wire experience must be screened off from the perceiving subject if she is to be prevented from recognising the object of her perception from the structure of the experience she is undergoing. The perspectival structure of the redirected wire experience cannot be reflected in the described phenomenological content of the experience — rather, it must be built into descriptions of the

cases that perceptions had through that faculty are experienced by the perceiving subject as bearing just the same sort of non-perspectival structure as her ordinary episodes of internal bodily awareness.

To introduce such a systematic illusion of perspective (or, more precisely, of a *lack* of perspective) into descriptions of redirected wire cases, however, is to concede that the epistemic grounds underlying internal bodily awareness judgments are not the same as those grounding judgments of redirected wire perception. Whether or not the identity assumption can survive structural differences, it is certain that it cannot survive systematic differences in veridicality. The identity assumption, if the phenomenological indistinguishability condition is to be met, must be rejected.

None of this, of course, rules out the counterfactual possibility of a sense modality similar to internal bodily awareness in many respects, but that is perspectivally structured. It is, moreover, perfectly conceivable that such a modality could enable perceptions of a body other than the perceiver's own 'from the inside'. In this sense, the redirected wire proponent can maintain, there is no problem with the conceivability of redirected wire cases. To understand redirected wire cases as involving a distinct, albeit similar, sense modality to internal bodily awareness, however, is to give up on the crucial identity assumption on which the threat of redirected wire cases crucially rests. The conceivability of redirected wire cases now shows us nothing at all about the identification-involvement of ordinary internal bodily awareness; it demonstrates only that there is a conceivable nearby sense modality through which other bodies could be perceived 'from the inside', a modality whose identification-involvement has no bearing on our understanding of the structure of internal bodily awareness.

There are two findings to this discussion on redirected wires cases, one methodological, the other substantive.

The first has to do with a point often taken for granted by proponents of the cases — namely, the assumption that redirected wire thought experiments can be taken to meet the necessary conditions for effectiveness by mere stipulation alone. Armstrong, for instance, simply mandates redirected wire cases as those in which we 'become aware e.g. of another's limbs, in much the same sort of

way that we become aware of the motion of our own limbs' (Armstrong 1984, p. 113). The first finding of the above discussion, is that such a result is not so easily won; a little more careful fingerwork is required in filling out the details of the cases if the central conditions on their efficacy are to be shown to be met.

The second finding is that these conditions *cannot* be met. Unlike vision, we saw, internal bodily awareness is non-perspectivally structured; there is no separate viewpoint from which internal perceptions of bodily states and events are had. Without such a separation between a perceptual point of origin and the object perceived, however, the kind of perceptual displacement described by redirected wire cases looses its intelligibility — there is no suitable relation available between the perceiver and the perceived into which a new viewpoint could be imaginatively inserted. For a non-perspectival mode of self-perception, rather, a shift in the body perceived brings with it a shift in the perceiver.

Traditional redirected wire cases, then, pose no threat to *internal bodily awareness immunity*. Once the distinctive spatial structure of internal bodily awareness is built into full descriptions of the cases, they will either be too weak to do any damage to that claim, or so strong as to collapse into internal incoherence.

5.1.2 Somatoparaphrenia

A second counterexample to *internal bodily awareness immunity* might seem to be provided by the delusion of somatoparaphrenia. This would be a counterexample of an empirical kind. Somatoparaphrenia is a delusion, most frequently associated with neurological damage in the right parietal cortex, in which sufferers either deny ownership of one of their own limbs or, in extreme cases, attribute the limb to someone else. The delusion is commonly accompanied by anosognosia (lack of acknowledgement that one is undergoing a delusion) and unilateral neglect (lack of attention or awareness to the spatial region on the side of the body of the affected limb). One of the remarkable features of the condition is that sufferers retain the capacity to feel and report on sensations in the affected limb, even while denying that the limb is their own. The sufferer can even, as demonstrated by the following quote, come to feel rather

oppressed by the pain in the disowned limb:

> Once home could I ask my wife, from time to time, to remove this left arm and put it in the cupboard for a few hours in order to have some relief from pain? (Maravita 2008, p.102)

These, then, are cases in which a subject comes to know about how things are with her own body through internal bodily awareness, but in which she is mistaken about *who it is* that she comes to know about in this way.

Is this a genuine counterexample to *internal bodily awareness immunity*? It is not. This is for the same reason that cases of thought insertion do no damage to the claimed immunity to error through misidentification of self-ascriptive judgments made on the basis of introspection. That is, that what these cases demonstrate is the possibility of internal bodily awareness based false negative errors, in which a subject fails to self-ascribe a property that she in fact instantiates. As we saw in §4.1.1, however, claims to immunity to error through misidentification affirm the impossibility of false-positive errors of a certain kind; they state that a subject couldn't be wrong, on those grounds, *in self-ascribing* the property. There is no threat here to the immunity to error through misidentification of self-ascriptive judgments made on the basis of internal bodily awareness.[6]

5.2 Multimodal bodily awareness

The last section was concerned with internal bodily awareness. There is a sense, however, in which treating internal bodily awareness as an independent faculty is something of an idealisation. It is rare that we receive information through, e.g., proprioception or nociception in perfect isolation from perceptual input from the five traditional exteroceptive sense modalities of vision, olfaction, audition, touch and taste. Indeed, recent empirical work shows that a great majority of our perceptual experiences are ineliminably crossmodal, and even seemingly monomodal perceptual episodes are typically influenced in substantial, complicated and often unnoticed ways by the other senses. This

[6] De Vignemont makes the same point in (de Vignemont 2012)

is no less the case in self-perception; as de Vignemont writes, 'empirical findings have shown that the normal way of gaining bodily self-knowledge is not via the body senses per se but rather via the integration of body senses and vision. Bodily self-knowledge is primarily multimodal.' (de Vignemont 2011, p.241)

There are a number of different ways of filling out the claim that our self-perception is primarily, or characteristically, irreducibly multimodal.[7] One might, for instance, be making a claim about the neural processes or physical sensory systems involved in the production of such experiences. Although this will surely be part of the full story, the claim for our purposes is more closely tied to facts about what it is typically like to undergo perceptual experiences of our own bodies. The claim that bodily awareness is primarily multimodal as it is intended here is the claim that the phenomenological and representational content of the experience is determined by incoming perceptual information coming from a number of different internal and external modalities in a way that is not merely summative — multimodal bodily awareness delivers a unified and irreducibly multimodal experience as of a single bodily object.

Here is the hypothesis that multimodal bodily awareness is a form of self-perception that issues in judgments with immunity to error through misidentification relative to uses of the first person concept:

> *Multimodal bodily awareness immunity:* I cannot be wrong, in forming a judgment about an occurrent bodily state, event or process of the form *I am F* (or, *I am F-ing*)on the basis of multimodal bodily awareness solely in virtue of a misidentificatory mistake about whether or not it is *me* that I know to be *F* (*F-ing*) on those grounds.

There is a prima facie difference between the initial plausibility of *internal bodily awareness immunity* and that of *multimodal bodily awareness immunity*. The former thesis was an intuitively powerful one. It seemed right to think that internal bodily awareness is a way that we have of perceiving our own bodies alone. The burden in the last section, then, was with its opponents to show

[7] For an extremely useful systematic taxonomy of the various distinctions one might be drawing on making such claims, see (Macpherson 2011)

why this is not so. Self-perception through the exteroceptive sense modalities, on the other hand, is clearly not so epistemically secure. I can perfectly easily see, hear, touch, taste and smell other people; and with that possibility comes the corollary that I might *think* that I am so perceiving myself when I am in fact perceiving another. Now multimodal bodily awareness is a form of self-perception that combines together these two clusters of sense modalities with their respective epistemic profiles, and it is not at all obvious what effect this integration will have on the epistemic status of the resulting unified experiential state. Will the epistemic robustness of internal bodily awareness carry over into the combined state, or will that combined state be contaminated by the epistemic vulnerability of the external senses?

I will argue that multimodal bodily awareness does preserve the epistemic privileges of internal bodily awareness. But first, here are three kinds of illusion that would seem to show that it does not.

5.2.1 Three illusions

Each of the three experimental paradigms in this section work by artificial manipulation of inputs into the processing system by which exteroceptive and somatosensory signals from our own bodies and its parts are normally integrated. There is, with respect to this integrative mechanism, a distinct hierarchy of senses. Where the information encoded by simulataneously incoming intersensory signals contains inconsistencies, the integration process appears to favour information derived from some sense modalities (vision, touch) over others. The result is a unified but distorted multimodal experience, that exaggerates information points coming in through the dominant senses, and minimises assimilation of any conflicting information from the recessive modalities.

Rubber hand illusion

In the rubber hand illusion, subjects are asked to sit with one of their forearms resting on a table in front of them. The arm is visually screened off from the subject and a rubber model hand is placed on the table directly in front of them,

in full view. An experimenter then uses two small paintbrushes to stroke, with synchronous movements, both the unseen real hand and the seen rubber hand. After a few minutes in this condition most subjects report undergoing an illusory experience often described using language of ownership — "I found myself looking at the dummy hand thinking it was actually my own." (Botvinick and Cohen 1998). More specifically, after ten minutes of stroking, the ten experimental subjects in the original experiment by Botvinick and Cohen all either agreed or strongly agreed with the following three statements:

> It seemed as if I were feeling the touch of the paintbrush in the location where I saw the rubber hand touched.
>
> It seemed as though the touch I felt was caused by the paintbrush touching the rubber hand.
>
> I felt as if the rubber hand were my hand.
>
> (Botvinick and Cohen 1998)

The experimenters hypothesised that the illusion was the product of intersensory bias towards vision — where there is spatial inconsistency in received crossmodal sensory signals, the suggestion is, the relevant integrative neural system prioritises visual information. The subjects come to feel their hand to be at the position they see the rubber hand to be. We might think that this visual prejudice is hardly surprising from an evolutionary perspective. De Vignemont explains:

> The major influence of vision on bodily experiences is not accidental. It has a strong adaptive value. On the one hand, most of the time vision is more reliable than proprioception at determining spatial information. Combining visual information to proprioceptive information thus increases the accuracy of bodily self-knowledge. On the other hand, our body navigates in and interacts with the external world, which is given to us mainly through vision.(de Vignemont 2011, p. 242)

The rubber hand illusion presents a prima facie challenge to *multimodal bodily awareness immunity*. If the illusory effect is sufficiently strong, we must allow for the possibility that the experiencing subject might go on to a form false

judgment, that *my hand is there* (where 'there' refers to the position of the rubber hand). The subject will then have gone wrong in forming a first person judgment on the basis of multimodal bodily awareness, solely in virtue of a misidentificatory error about whose hand it is to which they have epistemic access on those grounds.

Body transfer illusion

The body transfer illusion paradigm is similar in design to the rubber hand illusion — indeed, was initially inspired by it[8] — but serves to disrupt the subject's experience of the position of her entire body rather than just of one of its parts. The subject wears a headset fully covering her eyes. A video is is then shown on a screen inside the headset of the back of her own body from two metres away being stroked rhythmically. The experimenter then strokes the subject's back in synchrony with the strokes seen on the video. What the subject sees, then, is an image of the back of her own body two metres in front of her being stroked, and what she feels is the simultaneous haptic sensations of stroking on her back. After a short time the subject begins to feel as if her body is located at the position at which she seems to see the body in the projected image. Her entire body feels displaced to a position a few metres in front of her. This illusory effect is again explained by (and demonstrates the astonishing force of) the dominance of visual information in the intersensory integration process that feeds into our multimodal experiences of our bodies.

To show up as an apparent counterexample to *multimodal bodily awareness immunity* we need to consider a slight variation on this basic paradigm. That is because, slightly misleadingly, the apparent misidentification in the original case is with the subject herself. Bigna Lenggenhager and her colleagues found that the same effect can be induced by replacing the video of the subject's own back with footage of a mannequin being stroked where, again, the stroking is synchronous with the experimenter's stroking of their own back. The effect is striking; as Lengennhager et al write, '[o]ur results show that humans systematically experience a virtual body as if it were their own when visually

[8] See (Bigna Lenggenhager 2007, p.2)

presented in their anterior extra-personal space and stroked synchronously.' (Bigna Lenggenhager 2007) We can suppose that, taking such an experience at face value, a subject might come to form a judgment that is in error through a misidentification. Suppose, for instance, that she comes to judge that *I am over there*, on the basis of multimodal bodily awareness. Her judgment would be mistaken, and it would be mistaken through a misidentification. She would be wrong solely in virtue of a misidentificatory mistake about whether or not it's *she herself* that she knows to be over there on those grounds; it's not her but the perceived mannequin who is presented as being in that location.[9]

Nose displacement illusion

The third illusion is again structurally similar, but shows that the distortion needn't come from vision — it comes, in this case, from touch. A blindfolded experimental subject sits in a chair with an 'accomplice' standing in front of him, and the experimenter to one side. The experimenter takes the subject's finger and uses it to tap the accomplice's nose with an irregular pattern. At the same time, the experimenter uses their own finger to synchronously tap the subject's nose. As Ramachandran and his colleagues put the result:

> After a few seconds of this procedure, the subject develops the uncanny illusion that his nose has either been dislocated, or has been stretched out several feed forwards, demonstrating the striking plasticity or malleability of our body image. (Ramachandran and Hirstein 1998, p.1622)

The more irregular the tapping rhythm, the stronger the illusion.

A plausible explanation of this illusion stays close to the one given in the first two experiments we have seen, with the only difference being that in this case the conflict is between the spatial information encoded by the subject's (outwardly directed) tactile sensations and by his (inwardly directed) haptic sensations. The bias in this case is with touch; the subject comes to feel his nose

[9]There is a slight complication with this case, in that the judgment would not be *solely* in error through a misidentification; it would also be mistaken because there is, in fact, nothing there at all. It's at least plausible, however, that the same effect might be generated without a headset, where the subject sees a real mannequin being stroked two metres in front of her. The judgment, in this case, *would* seem to be a genuine case of error solely through misidentification.

to be at the location of the place that he feels the nose he is touching with his finger to be.

Once again, this is a potential counterexample to *multimodal bodily awareness immunity.* Were the subject to come to form the judgment on the grounds of this multimodal experience that *my nose is out there*, he would be mistaken solely through a mistake of identification. It's right that he knows, on those grounds, that *somebody's* nose is out there, but he has gone wrong in identifying that somebody as himself.

5.2.2 The case for *multimodal bodily awareness immunity*

These three illusions, and others like them, might be taken to seriously dent any starting plausibility to *multimodal bodily awareness immunity.* After all, they seem to be real-life cases in which subjects risk forming first person judgments on the basis of multimodal bodily awareness that are in error through misidentification relative to uses of the first person concept.

I think we nevertheless have reason to uphold it. The argument, in outline, is this. We have already seen that internal bodily awareness is a source of judgments with immunity to error through misidentification relative to uses of the first person concept. This is the first premise of the argument, defended in §5.1. The second premise — to be established in this section — is that multimodal bodily awareness is the result of a non-inferential, and so non-identity involving, integrative process of incoming crossmodal information. The conclusion is that multimodal bodily awareness is likewise a source of judgments with immunity to error through misidentification relative to uses of the first person concept. This is because any possibility of a *mis*identification supervenes on the presence of an *identification* in the judgment's epistemic structure. If the judgment's formation lacks an identificatory structure, then there can be no possibility that it could be in error through a misidentification.[10]

[10]A very similar argument is given in (de Vignemont 2011); 'In a nutshell, (i) only signals assigned to a common source are integrated together, (ii) the assignment to a common source results from a subpersonal comparative process that does not rely on self-identification, and (iii) proprioceptive signals are only from one's own body. Combined together, these three principles guarantee that bodily IEM is preserved in integration-based bodily self-ascriptions.' (de Vignemont 2011, p.239)

Before turning to this argument, a quick parenthetical remark on an aspect of the debate surrounding the notion of immunity to error through misidentification that I have not yet addressed, but that might be prompted by this way of putting the argument of this section. Some writers — e.g. Coliva and Wright — have argued that it's not only that the judgment's *grounds* need to be identification-free; there must also be no background *presupposition* of identification if the final judgment is to be immune to error through misidentification. [11] Presuppositions, on these views, are no part of what lead a subject to form the judgment in the first place, but make up the background assumptions taken for granted in the context in which the judgment is formed:

> The information, while neither entertained in a process of conscious inference, nor part of the subject's own rational grounds — part of what he would appeal to in order to justify his final judgment — may be such that, were it somehow to fall into question a (rational and appropriately equipped) subject would be prepared to withdraw from teh judgment that *P*. In such a case, I shall say that the information provides certain *background presuppositions* on which the subject's final judgment rests. (Coliva 2006, p. 416)

Presuppositions, then, are answerable not to considerations of judgment-formation, but to a test of posthoc retraction; is there an identity assumption in the background that could force a rescindment on the final judgment if called to the thinker's attention and brought into doubt?

I have no ambitions here to settle whether or not this is the right account of immunity to error through misidentification — that is a substantive question that is only orthogonally related to the present discussion. I only want to point out that the arguments of this section would work against this 'presupposition view' of immunity to error through misidentification just as much as against the more standard 'formative grounds view'. That is because if it's right that multimodal experience is fully integrated in the way that will be argued for in this section, then it is an epistemic channel that cannot be reduced to its component monomodal contributions. And if *that's* right, then there is

[11] (Wright 2012), (Coliva 2006)

no candidate identity belief between monomodally perceived bodies, or body parts, that could force a retraction of the multimodally-based first person judgment. Since multimodal experience is irreducible, any appeal to monomodal perceptions would be a shift in the epistemic grounds; doubt that the body part I monomodally see is the same as the body part I monomodally proprioceive does not amount to doubt that the body part I multimodally perceive through vision and proprioception is one and the same.[12]

The task now, then, is to defend the second premise — to say why we should accept that the process by which incoming sensory information of different kinds is integrated into a unified conscious experience should be understood as an epistemically non-inferential process.[13] Support for this premise comes primarily in the form of reasons to reject an inferential understanding of the integrative process.

What reason do we have to resist the idea that multimodal integration is identity-belief involving, and so inferential? We might begin by noticing that a characterisation is owed of the inference that could account for a multimodal experience as of a single bodily object instantiating multiple distinct property-kinds. This amounts to something like a challenge to an inferentialist response to the so-called binding problem — or, at least, to a version of that problem as applied to the case of crossmodal self-perception. What, the question is, could move me to infer that a single bodily object is the bearer of all these different properties? Why would I infer that that the reddish arm that I see from the outside is the same object as the painful body-part I feel from the inside? Here are two answers that cannot be used by the proponent of the inferential model of multimodal self-perception.

Consider a purely visual version of the same question: what could move me to infer that the red object I see is the same as the round shape I see? A promising response might be the spatio-temporal coincidence of the redness

[12] Of course, it might cause the thinker to doubt *whether* she is having a genuine multimodal experience of her body, or two separate monomodal perceptions, but that is another question. Such a case would be a challenge to the *transparent* immunity to error through misidentification of multimodal bodily judgments, but not to their immunity to error through misidentification simpliciter (see pp.67-8).

[13] All sides will be in agreement, of course, that the process is *psychologically* non-inferential.

and roundness — e.g. Campbell, '[a]ttention to a particular location is what binds together the features at that location as features of a single object.' (Campbell 1999a, p.98) This answer is made available in this visual case by the fact that the spatial positions of the property instantiations are presented to the perceiver relative to a single spatial frame of reference. This means that spatio-temporal colocation is immediately recognisable. Plausibly, however, this is not also true of a multimodal case in which sensory signals are received from both exteroceptive and interoceptive modalities; in Merleau-Ponty's memorable phrase, '[t]he outline of my body is a frontier which ordinary spatial relations do not cross.' (Merleau-Ponty 1989, p.98) A piece of commonplace evidence for this claim is that we are notably bad at determining the colocation of internally and externally perceived states and events; finding the bump on one's head that corresponds to an internal point of felt soreness is much more often a matter of tactile exploration than one of effortless recognition. It seems that the inferentialist cannot appeal directly to the sameness of spatial frame of reference to solve the binding problem facing the inferential model of multimodal self-perception.

Another unavailable answer comes from Evans. Evans's answer to Molyneaux's question — the question of whether a congenitally blind subject could immediately visually recognise a square upon the recovery of sight based only on her past tactile experiences of squares — is that she could. This, he argues, is because the spatial information encoded by vision and touch are necessarily relativised to the same spatial framework, because they carry contents with overlapping implications for action. The univocity of this egocentric 'behavioural space' (Evans 1985, p.386) in which both visually and tactiley perceived squares are experienced ensures the univocity of the shape concept possessed by both congenitally blind and sighted subjects. Similarly, we might think that our status as agents acting in a single external spatial framework is one way of solving the binding problem in the purely visual version of the case. I see the redness and the roundness as colocated (and so bound together in a single object) because the redness and the roundness are both experienced as associated with affordances in a single egocentric behavioural space. But the same cannot be said in the multimodal version of the challenge in which

sensory input is derived from both interoceptive and exteroceptive modalities. We do not *act* in internal space in anything like the same way that we act in our external environments, and to the extent that we do have agential control over happenings in our internal body space (holding one's breath, tensing a muscle) the role played by spatial information about the target body-part is very different to the role played by spatial information in Evans' behavioural space: there are no tracking requirements in internal body space, we do not need to locate a muscle before tensing it in the way that we would need to locate a rattle heard in the dark before reaching for it. The inferentialist cannot appeal straightforwardly to action affordances to secure the coincidence of spatial frame of reference between the internal and the external sense modalities.

Together, these considerations tell against a view on which the integrative process by which we synthesise incoming sensory signals from multiple modalities into a unified and irreducibly crossmodal experience as of a single body, or body-part, is to be understood as inferential. Without a better response to the binding problem, the inferentialist is left without an account of what could possibly get such an inference off the ground.

We have reason to think, rather, that multimodal integration occurs at a much earlier stage of information processing than the inferential model would suggest. As de Vignemont explains:

> [M]ultimodal integration occurs very early on in the perceptual process, at a stage where raw modality-specific sensory signals are not available to the subject. [...] [T]here is no need to first identify the source to determine that the sensory signals come from the same source. It suffices to compare the sensory signals themselves and the information they carry. [...] Hence, the visual system does not have to identify the seen body as one's own body. Rather, the properties of the seen body are compared with the properties of the felt body (e.g., location, posture), and if they are similar enough, the visual and proprioceptive signals are melted into a multimodal perceptual experience of one's hand. (de Vignemont 2011, p.244)

On this picture, the separate strands of monomodal sensory input are already combined into a unified experiential state prior to the its becoming available for any personal-level reasoning like inference.

Given, then, that the possibility of a misidentification error depends on the presence of an identification, we can conclude that the privileged status of internal bodily awareness is preserved when combined into states of multimodal bodily awareness, understood on this early-processing model. We have reason, that is to say, to endorse *multimodal bodily awareness immunity*.

5.2.3 Back to the challenge from the three illusions

We still face the challenge posed by the three illusions in §5.2.1. We now have reason to think that judgments formed on the basis of multimodal bodily awareness will be immune to error through misidentification relative to uses of the first person concept. What, then, are we to make of the data thrown up by these illusions, that seem to show that they are not?

One response to this challenge is best illustrated by considering a variation of the rubber hand illusion. Erhsson and his colleagues (Erhsson 2005) designed a non-visual version of the experiment. The subjects, blindfolded this time, are again asked to rest one arm on a table in front of them, and a rubber hand is placed on the table beside it. The experimenter then uses the subject's second hand to touch the rubber hand, while simultaneously touching the subject's resting hand with synchronous touches. After a while, subjects report feeling a proprioceptive drift in the direction of the rubber hand, and an illusory experience as of touching their own hand. These are of course interesting results in themselves, but what is crucial for our purposes is the fact that when asked to point to their resting hand, the subjects point neither to their own hand nor to the rubber hand, but to somewhere in between the two. This is crucial because it shows that the mistake these subjects are making in the grip of this amended rubber hand illusion is, at least in the first instance, a mistake of *mislocalisation* — an error for sure, but not an error of misidentification.

Out of this observation emerges a new reading of the respective mistakes being made in the three illusions of §5.2.1, a reading that leaves intact the argued immunity to error through misidentification of multimodal bodily awareness based judgments. There are, the suggestion is, not one but two mistakes being made in each of these cases: an initial error of mislocalisation made on the

basis of multimodal bodily awareness alone, and a subsequent confabulatory error of misidentification based on the visual or tactile identification of a likely candidate source of the distorted experience.

Take, first, the original rubber hand illusion. The present suggestion is that the illusion operates by a two-step process. It works, in the first instance, by an interference of the spatial content of the subject's crossmodal experience; she still perceives her own hand through multimodal bodily awareness, but the spatial misinformation that is fed into the multimodal state causes her to misperceive its location. The second step is confabulatory, and is based on vision alone. The subject sees the rubber hand positioned at or near the location at which she misperceives her hand, and goes on to identify it as a likely source of her spatially distorted multimodal experience. This case is to be treated, that is to say, much like the amended rubber hand case of two paragraphs ago, except that this time there is a likely-looking candidate around upon which to alight in making sense of the distorted experience. The multimodal experience itself is not vulnerable to errors of misidentification, but the state that results from combining it with an additional vison-based identification is.

Similar stories can be given for the body transfer and nose displacement illusions. In both cases the first mistake, and the only mistake made on the basis of the integrated multimodal experience alone, is one of mislocation — in the case of the body transfer illusion the subject mislocates her entire body, in the nose displacement illusion, the position or dimensions of her nose. In both cases, however, there is an available candidate around to be identified as the source of the distorted experience (the projected mannequin image, the touched nose). In both cases, the error that is liable to be made on the basis of the integrated multimodal experience alone is one of mislocation, but the formation of new epistemic grounds out of a combination of that experience with a vision- or touch-based identification renders the final state open to errors of misidentification. So understood, these illusions pose no threat to *multimodal bodily awareness immunity*.

5.3 Conclusion

The aim of these two chapters has been to identify some of the forms of self-awareness that play the central role in a full account of first person thought that has been carved out over chapters 1-3 — a special cluster of forms of self-awareness that will be marked by the immunity to error through misidentification relative to uses of the first person concept of judgments formed on their basis. I have set out and defended this status in the case of introspection and episodic memory in chapter 4, and internal bodily awareness and multimodal bodily awareness in this chapter. The piecemeal strategy taken up in these chapters is unavoidably exploratory and inexhaustive; there are no doubt many others I have failed to consider.[14]

In the next chapter I turn from first to second person thought.

[14]Two such further faculties are our capacity for visual self-location, and the distinctive way in which we know about our own actions.

Chapter 6

From me to you

There is a special kind of interconnectedness between the thoughts canonically expressed in English by uses of 'I' and those canonically expressed by 'you'. This interconnectedness shows up in our patterns of understanding when it comes to related utterances: in order to understand what you are saying when you express a thought about me using the word 'you', I must think an 'I'-thought to myself. As McDowell writes:

> Suppose someone says to me, "You have mud on your face". If I am to understand him, I must entertain an 'I'-thought, thinking something to this effect: "I have mud on my face: that is what he is saying". (McDowell 1984, p. 291)

A pair of questions immediately follow from this observation. First, what kind of thought is canonically expressed by uses of the second person pronoun, and in particular, is there a *distinctive* kind of thought that is so expressed? And second, what is the relation between the kind of thought so characterised and first person thought such that these McDowellian understanding constraints are to be accounted for?

Answers to these questions must hang together. A number of recent writers, including Sebastian Rödl, Michael Thompson, José Louis Bermudez, and Guy Longworth, have answered the first with the claim that there is no such thing as distinctively second personal thought; second person utterances express thoughts of the very same kind as those containing sincere uses of the

first person pronoun.[1] As Rödl explains,

> "You..." said by me to you and "I..." said by you in taking up the address, express the same act of thinking, they express the same thought. Therefore it is wrong to oppose second person thought to first person thought. This is a difference in the means of expression, not in the thought expressed.(Rödl 2007, p. 197)

This answer to the first question supplies a straightforward response to the second. Given that second person thought is nothing more than first person thought, it is no surprise that understanding a second person utterance involves entertaining a first person thought. This is just to say that understanding the utterance involves entertaining the thought expressed. The thought grasped by speaker and hearer are the same, only the canonical modes of expression from the two perspectives differ.

To say that second person thought is the same as first person thought is not the only way to be a reductionist about second person thought. Others deny that there are such strict constraints on the kinds of thought expressible by 'you', but take such utterances to be typically expressive of a demonstrative way of thinking of the other (Richard Heck), or reduce apparent uses of a second person concept in thought to complexes of first and third person thought (Christopher Peacocke).[2] The aim of this chapter is to argue against reductionists of all stripes that there *is* such thing as distinctive second person thought. This account, of course, will need to be accompanied by a complementary account of the relation between such thought and first person thought, such that the above McDowellian understanding constraints can be accommodated.

In §6.1 I will identify two arguments in the literature on reductionism about second person thought. I disarm the first, but defer response to the second and the strongest — the argument from addressing — to a later section. I motivate and present my account of second person thought in §6.2. In the third section

[1] See (Rödl 2007), (Thompson 2012), (Bermudez 2005), (Longworth 2014), (Longworth 2013), though it should be noted that Longworth only argues that *some* second person thoughts are the same as first person thoughts; he takes others to be identical with 'that person' thoughts.

[2] See (Heck 2002) and (Peacocke 2014)

I return to the argument from addressing, and show that it need not undermine the account of second person thought just given. I consider three kinds of apparent counterexamples in §6.4. I then turn, in §6.5, to the question of the relation between second and first person thought in light of the McDowellian understanding constraints of this introduction. I argue that the account of second person thought developed in this chapter accommodates those constraints just as naturally as its opponents.

6.1 Two arguments for reductionism

What kinds of consideration have moved reductionists about second person thought? One sort of consideration for a view that identifies second and first person thought, which I mention here only to put to one side, will be of interest only to those pursuing a particular neo-Fregean project: namely, that of developing Frege's account of first person thought in a way that coheres with other features of his theory. In 'Sharing Thoughts about Oneself' (2013), for instance, Longworth defends the compatibility of an identity claim between first and second person thought (or, as he terms it, the *shareability thesis* about first person thought) with Frege's criterion of thought individuation.[3] John Perry, and more recently José Louis Bermudez, have argued that a denial of that identity claim (or, of the shareability of first person thought) would stand in intolerable tension with Frege's ontology of thoughts, and in particular with their postulated objectivity. It would be at odds, they argue, with what Perry describes as 'Frege's timeless realm of generally accessible thoughts'.[4] These sorts of considerations can be left aside for the purposes of the present project, which do not include finding a way to render Frege's theory internally consistent.

Instead, I will focus on two self-standing arguments from the literature.

[3](Longworth 2013)

[4](Perry 1977, p. 492),(Bermudez 2005); see Gareth Evans's (Evans 1981) as a response to this point in Perry, and Daniel Morgan's (Morgan 2009) as a response to this point in Bermudez.

6.1.1 The argument from the naïve view of communication

One of the most prevalent arguments for the identification of first and second person thought — and so against a non-reductionist account of the latter — takes its shape from the attempt to understand the McDowellian understanding constraints described in the introduction. Entry points and emphases differ between different presentations, but we can take a clear statement of the basic argument from Longworth.[5] After setting out the above passage from McDowell, he asks, 'why should it be that understanding someone who makes use of "You" requires that one think a self-conscious thought about oneself?'.

The answer he gives is that the thought expressed with the speaker's use of 'you' *just is* the self-conscious thought one thinks in understanding it:

> A natural hypothesis would be that that is so only because one is thereby thinking a thought that they expressed and, moreover, because the thought that they expressed was one that they also entertained. For if things were not like that — if, for example, the thought about one that they entertained were distinct from any thought that one would express by the use of "I" — then it would be difficult to see why it should be required of one, in order to understand them, that one entertain a thought that one would express by the use of "I". (p.298)

In this, what I take to be its most compelling, form, the argument is of abductive force.[6] The best explanation of the fact that in order to understand what has been said by a second person utterance (directed towards me), I must think a first person thought, is that the thought thereby entertained and the thought thereby expressed are one and the same. In Rödl's unambiguous phrase, 'Second person thought is first person thought' ((Rödl 2007, p. 197)). And if second person thought is first person thought, of course, then there is no such thing as distinctive, or irreducible, second person thought.

[5] See (Longworth 2014, pp. 12–13); (Bermudez 2005, pp. 184–5); (McDowell 1984, pp. 289-290); (Rödl 2007, p. 196); (Peacocke 2014, pp. 248–9) for formulations, but not necessarily endorsements, of the argument. The most common entry point is, in fact, not to do with second person thought directly, but a question about how first person thoughts could be communicable on a Fregean view.

[6] Some others, such as Rödl, seem to give it a stronger rendering than this; the problems I level against the argument here would be equally applicable against a stronger formulation.

The problem with this argument is that it draws heavily on a questionable background view of communication. The identification of the thought entertained with the thought expressed best explains McDowell's observation only on the assumption that understanding (or, at least, the kind of understanding required for communication) requires the sharing of a single thought between speaker and hearer. This is the view Heck calls *the naïve view of communication* — the view that, as he puts it, 'what my words mean is precisely what I already believe and you come to believe: when you grasp the content of my assertion, you thereby grasp the very Thought I believe and am trying to communicate to you.' (p.6)

For Heck, the assumption of this view is indefensible. Sharing of the very same thought *couldn't* be a requirement on communication, he argues, given that we manifestly do sometimes communicate successfully in situations in which the same thought is unavailable to both speaker and hearer. Heck gives two such examples. The first case, and the one to which he devotes most attention, is that of first person thought:

> The belief that someone expresses when she says "I am a philosopher" is the self-conscious belief that she herself is a philosopher. But the belief I form, if I accept what she says as true, is not the self-conscious belief that she is a philosopher: I cannot so much as entertain that belief. [...] The belief I acquire [...] is, instead, the belief that she (the speaker) is a philosopher, a belief that involves a demonstrative (and not a self-conscious) way of thinking of her. (Heck 2002, pp. 20–21)

In the present context this example is of little help; supporters of the above argument would presumably simply reject it outright. After all, that the belief (or thought) that the speaker expresses when she says 'I am a philosopher' is unshareable is precisely what they deny.

Heck also, however, appeals to a second case, which at least initially seems a little more promising, at least for our purposes. It involves a pair of interlocutors discussing a bottle perceived from distinct perspectives. The two subjects, he insists, need not be thinking about the bottle in exactly the same way, and indeed 'my perspective on the bottle may be sufficiently different from hers

that my Thought is, by the usual Fregean criterion, different from the one the speaker was expressing.' (Heck 2002, p. 21) Clearly, Heck urges, this would not compel us to the verdict that communication between them about the bottle is impossible.

Imogen Dickie and Gurpreet Rattan have convincingly argued for a charge against Heck of mishandling this case. So long as the subjects are in a position to communicate about the bottle, they submit, the subjects *are* thinking about the bottle in the same way; 'speaker and hearer are able to understand one another's uses of the demonstrative in virtue of the fact that they are jointly attending to its referent.'(Dickie and Rattan 2010, p. 137). The fact that there can also be cases of demonstrative thought that fall short of the Fregean criterion for thought individuation should not be read into the present case.

These cases aside, however, Heck's discussion highlights a crucial, if modest, point, which is that the conception of communication drawn on by the argument is not mandatory. There are other options. This is also a point made by McDowell:

> Frege's troubles about 'I' [...] result, rather, from the assumption [...] that communication must involve a sharing of thoughts between communicator and audience. That assumption is quite natural, and Frege seems to take it for granted. But there is no obvious reason why he could not have held, instead, that in linguistic interchange of the appropriate kind, mutual understanding — which is what successful communication achieves — requires not shared thoughts but different thoughts which, however, stand and are mutually known to stand in a suitable relation of correspondence.(McDowell 1984, p. 290)

McDowell says little more than this on the point, but he leaves us with the inviting suggestion that the relation between the thoughts entertained by a speaker and her hearer respectively in a communicative exchange need not be as strong as identity.[7] Indeed, not only does this idea that they may stand in some weaker relation of correspondence demonstrate the presence of an alternative theoretical option — it also has considerable initial appeal in its own right. Beyond the

[7]Though we might not want to go along with McDowell all the way in also adding a common knowledge constraint.

philosopher's fiction, it seems wildly implausible that many (any?) of the conversations in which we participate fit the model recommended by the naïve conception of communication, according to which complete thoughts are passed around, intact and unmodified, between participants. Conversations, rather, are messy affairs. The thoughts entertained, expressed, understood and responded to by interlocutors are surely far more elastically related to one another than this without any loss of successful communication. We might even think that if the bar on successful communication is set at the sharing of a single preserved thought between speaker and hearer, then it will turn out that we are radically less competent at communication than we might think — very few of our actual conversations would seem to be be candidates for passing this test.[8]

These sketchy remarks are, of course, no knock-down argument against the naïve view of communication. But they hopefully serve to animate the point that there is available a natural looking alternative conception of communication, a conception that would disarm the above argument for the identification of 'I'- and 'you'-thoughts. As things stand, then, there is not yet any reason to accept the argument's key background assumption, and so no reason to accept its concluding identity claim, nor that claim's reductionist corollary.

This rejection of the argument from the naïve view of communication incurs an explanatory debt. The argument provided resources with which to explain the McDowellian understanding constraints of the introduction. With the renunciation of the argument, we also turn down these explanatory resources. I will return to this debt in the final section of the chapter.

6.1.2 The argument from addressing

The second argument comes to view in some suggestive but persuasive remarks from Heck:

> [A]n utterance of 'you' refers to the person addressed in that utterance. [...] The phenomenon of the second-person is a linguistic one, bound up with the fact that utterances, as we make them, are typically directed to people,

[8] See (Recanati 2012b, p. 210), who also cites (Davies 1982), (Egan 2007) and (Ninan 2010), for further discussions of the idea that this model of communication, on which speaker and hearer must share a single thought, must be revised to accommodate indexical language.

> not just made to the cosmos. (If there were speakers of a language who never directed their utterances to their fellows, they would have no use for the second-person.) (Heck 2002, p. 12)

The second person pronoun, the idea is, is a linguistic device used to refer to the addressee of an utterance. To address someone, however, is a purely linguistic phenomenon; it is something that happens — as Michael Thompson puts it — only 'in language, in the noise, in the outward show of things', and not 'in the secret depths of the soul' (Thompson 2012).

There is, of course, a superficial sense in which the claim that addressing is a purely linguistic phenomenon is manifestly false. Upon observing a fellow shopper chose a neighbouring queue I might, somewhat smugly, savour the thought *I'm going to beat you*. Isn't this a way of addressing someone else in thought? It is, but not in the sense targeted by Heck. In this case I am treating the other person as an imagined interlocutor — we might even say that I am enacting a kind of (very quiet) way of talking to her that takes place entirely in my head. In characterising the notion of addressing as a purely linguistic phenomenon, Heck is not denying the possibility of internal articulations of second personal utterances. What he is denying is that there is a non-derivative correlate in thought of the linguistic act of addressing someone.

There is an obvious reason to think that Heck is right. Successfully addressing someone, at least at first pass, seems to require that they notice (or, perhaps, that there is at least a *chance* of them noticing) that the utterance is directed towards them. Without that, one's act would be nothing more than talking *in their direction*, or talking *at* them, acts both falling short of addressing them. But for that, of course, there must be some outwardly recognisable signal associated with the utterance to indicate that the utterance is directed towards them. There could be no analogue of this in the private domain of thought — at least, not for the non-telephathic creatures that we are. If second person thought is to march in step with its mode of linguistic expression, however, then its full characterisation would likewise need to advert to the notion of addressing. So long as this is right, it seems that there could be no such thing as distinctive second person thought.

This short argument seems to provide a strong prima facie consideration on the side of reductionism. Before dealing with it, it will be helpful to have an idea of what a non-reductionist picture of second person thought might look like.

6.2 A non-reductionist picture of second person thought

The argument from addressing is striking. It is striking because prior to encountering it (and, perhaps, other arguments for reductionism), it is overwhelmingly natural to suppose that one's second person utterances are expressive of second person thought. For the most part, we assume that our ways of talking track our ways of thinking; why should the second person be any different?[9] Insofar as reductionism amounts to a form of exceptionalism about the second person, then, we have reason to consider carefully the case on the other side.

The appeal of non-reductionism about second person thought begins from the idea that there is a reason that I use 'you' to refer to you, when I do. My use of the second person pronoun signals a special kind of relation holding between us. To see what this relation is, consider the differences in context that must hold if I am to think a thought about you canonically expressed with 'you', and one canonically expressed with 'that person'. In both cases, I must be aware of you. This is hardly surprising; they are both context-dependent thoughts I have about you. Beyond this point, however, things start to look rather different. For in order to express a thought about you using 'you', but not in in order to express one using 'that person', it seems that there also has to be awareness going in the the other direction: *you* also have to be aware of *me*. This kind of reciprocity was part of what was involved in the notion addressing drawn on in the last section. You cannot notice that my utterance is directed towards you without also being aware of me.

Even this, however, seems not yet to fully capture the awareness relations signalled by my use of 'you'. To see this, imagine that you and I are interested in, but a little shy of, one another. I am attending to you out of the corner of

[9] This isn't always true — it isn't true, for instance, of some purely grammatical features of language.

my eye, all the while making great pretence of being engaged in conversation with someone else, and you are doing likewise. We are, here, engaged in a state of mutual awareness — we are each attending to the other — but neither is aware of this fact. This does not yet seem like a context in which my thought about you will be apt for expression using a second person pronoun; we are still firmly in 'that person' terrain. If that is right, then what this case shows is that not only must I be aware of you, and you of me, but I must also be aware of featuring in your awareness if I am to express the thought I have about you using 'you'.[10]

Still, we might think, we have not gone far enough. Consider the following case, used by Lucy O'Brien to illustrate a slightly different, but neighbouring phenomenon she calls 'ordinary self-consciousness':

> Consider Hermione in Shakespeare's *Winter's Tale*. She stands, at the end of the play, taken to be a statue by those around her. Leontes, who falsely accused her of infidelity years before, comments on the statue saying "Hermione was not so much wrinkled, not so aged as this seems." (Act V, Scene 3). Hermione might surely feel self-conscious at his perusal, and embarrassed by his remark. (O'Brien 2011, p. 120)

The conditions listed so far have all been met. Hermione is aware of Leontes, Leontes is aware of Hermione, and Hermione is aware of Leontes' awareness of her. The scene, however, lacks symmetry; Leontes is not likewise conscious of Hermione's awareness of him. The question now is, do the described circumstances suffice for Hermione to think a thought about Leontes that she would naturally express using a second person pronoun? If the answer is yes, then the kind of thought naturally expressible with a second person pronoun would seem to require three levels of awareness: the thinker's awareness of the referent, the referent's awareness of the thinker, and the thinker's awareness of the referent's awareness of the thinker. I am going to argue that this is not

[10]Of course, with a bit of imaginative toying the shy thinker here might work herself into a position to think a *you*-thought of a kind; this sort of possibility is brought out more clearly with respect to the case that follows of Hermione (because less imaginative work is needed on the part of the thinker), but the remarks made there apply equally to the parallel possibility here.

enough, that a fourth level is needed — that the referent must also be aware of the thinker's awareness of the referent.

At first blush, however, it might seem evident the asymmetrical conditions described here *do* suffice for Hermione to think thoughts naturally expressible using second person pronouns. Indeed, it does not take much to imagine Hermione luxuriating in the entertainment of thoughts she might express as: 'You're one to talk — *you* were the one who gave me these wrinkles'. Clearly, her thoughts here would be most straightforwardly expressed — if express them she could — using sentences containing second person pronouns. Is this not enough to show that the asymmetrical conditions suffice for thought canonically expressible with the use of second person language? I think it is not.

The reason is that, far from insisting that these are thoughts whose natural expression would not take the shape of second person pronouns, it seems that these thoughts are *peculiarly tied* to their form of linguistic expression. Hermione, we might think, is imaginatively addressing her husband — she is gratifying herself by resentfully turning over in her mind the things she would say to him if only she could. But the closeness of these thoughts with their imagined expression is precisely what raises problems for appealing to them in the service of demonstrating that the present conditions suffice for the entertaining of a thought naturally expressed with 'you'. That is because what we have here, if this is the right reading of the case, is not really an autonomously entertained thought of a kind expressible through the use of a second person pronoun, but rather a phonetically repressed articulation of an imagined second person utterance, in the scope of which imagining the conditions might well look rather different. Hermione, the idea is, is *imagining* herself saying these things aloud to an attentive Leontes, who is aware of Hermione's awareness of him. Put another way, there is no easy way to isolate the conditions in a case of this kind to ensure that a fourth layer of awareness — the referent's awareness of the thinker's awareness of the referent — isn't being imaginatively hypothesized. We do not yet have a clean case showing the first three conditions of awareness to be enough.

It might be objected that this argument is too quick. Even if there is no easy way of ruling *out* that the entertaining of thoughts of this kind are only enabled

by the imagined positing of an extra fourth condition, we should not then rule that reading *in*. There might be an alternative understanding of Hermione's thought here that does not appeal to an imaginative context, and so that *would* show these first three conditions by themselves to be enough for Hermione to think a thought that would be naturally expressed with the use of a second person pronoun, if expressed at all.

There are interesting difficulties involved in the assessment of this counterfactual. In the closest world in which Hermione did express these thoughts, it is likely that the conditions of the situation would thereby be changed. The act of addressing Leontes aloud would create what Recanati has called *communication-specific facts* about the situation, or facts that do not exist independently of the communicative situation[11]. In this case, the relevant communication-specific fact would be that Leontes would be newly aware of Hermione's awareness of him. Perhaps, though, this difficulty can be circumvented by imagining a case in which Hermione expresses her thought, but not very loudly — perhaps she just mutters the relevant sentence under her breath. The case so imagined, the idea is, would reveal the above asymmetrical situation as one in which Hermione is able to think a thought naturally expressed with 'you', and so show that the first three conditions of awareness suffice for such a thought, without the addition of the fourth.

It's not clear, however, that even this manipulation of the case provides us with an uncluttered example of the kind of thought expressible with 'you' in perfect isolation from a posited fourth condition of awareness on the part of the referent. It's not clear, because it is not clear that Hermione is not once again building Leontes' imagined awareness of her awareness of him into the episode; plausibly she is only muttering under her breath what she would say aloud to him, if only she dared. One way of urging this interpretation is to press on the intuition her second person mutterings could be felicitously substituted with a third personal utterance ("he's one to talk — *he's* the one who gave me these wrinkles") with only a slight change in tone. Tellingly, the felicity of this interchangeability dissipates in a case in which Hermione really does pluck up the courage to address her second person utterance aloud to Leontes.

[11] See (Recanati 2012b, p.218)

What this shows is that the kind of thought expressed by a use of 'you' in fully symmetrical versions of the case — versions in which there is mutual awareness of mutual awareness — is distinct from the sort of thought entertainable before the fourth condition is introduced. This is so, even if it is possible for a thinker to entertain imagined internal articulations of the relevant kind prior to the actual obtaining of that fourth condition.

Adding a further layer of awareness on the part of Leontes to the scene, then, seems to make available to our protagonist thoughts expressible with 'you' that differ in important respects from those she has been in a position to entertain until now, thoughts that shake off their air of make-believe. Imagine now, for instance, that Hermione reveals herself to Leontes by a wink of the eye. Leontes' crashing realisation, we may imagine, will bring with it a wave of shame and horror. But the change brought about is not only with him. There has now been introduced onto the scene a kind of mutual recognition of mutual awareness, a change that has implications for the kinds of thought available to Hermione, as well as one with emotional and cognitive significance for Leontes. Hermione no longer needs to *imagine* herself to be interlocked in a kind of reciprocal conscious interaction with her husband. They now really are so related to one another.

The difference in the kind of thought now available to Hermione can be again brought out by third personal substitutability considerations. Earlier — before the introduction of Leontes' awareness of Hermione's awareness of him — Hermione's quietly muttered second personal utterance seemed to be felicitously replaceable with a third personal utterance, with the only difference levied being a subtle difference in tone. Can the same be said here? Intuitively, it seems not. Once the two are locked in mind-to-mind contact, there is a sincerity to Hermione's muttered use of the second person pronoun that blocks its straightforward substitution with a third person pronoun. This seems to be so, even if the muttering is too quiet for Leontes to hear. What we have here is a kind of thought expressible by a use of a second person pronoun, whose genuine availability seems to depend on the obtaining of the set of relations exemplified in this final version of the case of Hermione and Leontes.

Hermione's capacity for thought of this final kind, then, calls for four levels

of awareness. She must be aware of him, and he of her. She, moreover, must be aware of his awareness of her, and he her awareness of him. Taken altogether, this involves a phenomenon that Peacocke has discussed under the label *interpersonal self-consciousness*, 'a particular form of awareness that one features, oneself, in another person's consciousness, as a conscious subject' (Peacocke 2014, 236, abstract). In sincere and successful uses of the second person pronoun I am aware that I feature, myself, in your consciousness as a conscious subject, because I am aware that you are aware of me being aware of you. My use of the second person pronoun signals my interpersonal self-consciousness with respect to you.

Peacocke's formulation of interpersonal self-consciousness brings out an important point, which is that the specification of the iterating layers of self- and other-awareness cannot be neutral on the question of *how* we are each aware of the other. For Peacocke, I must be aware that I *myself, first personally* feature in your awareness *as a conscious subject.* The move is responsive to a worry raised by Thompson, who points out that so long as the awareness relations are specified in terms of variables — A is aware of B being aware of A's awareness of B — the formulation cannot suffice to capture common knowledge of the first order state of mutual awareness. (Thompson 2012) No matter how high one stacks these 'doxastic pancakes', Thompson explains, it remains possible that the parties do not know that it is *they* who are engaged in these awareness relations. It is not enough, then, to say as I did above that the distinctive cognitive significance of second person thought is that it signals a state of mutual awareness of mutual awareness holding between thinker and referent, or by filling out a characterisation of that state using variables. It must be stated, rather, as the condition that I am aware of you being aware of me (myself) being aware of you (yourself). We will return to this point in §5.5.

The route from here to a commitment to a distinctive kind of second person thought is not direct. That this is so is demonstrated by Peacocke himself, who agrees that interpersonal self-consciousness is essentially involved in the use and understanding of second person utterances:

> In a case of your successful communication with me in which you use the second person, I as audience know that you, the speaker, are aware that

> I know that you are saying that I am F. This is, in more than one way, an instance of ascriptive interpersonal self-consciousness on my part. (Peacocke 2014, p. 245)

In elaborating the nature of the states underlying such linguistic exchanges, however, he denies that we need appeal to a distinctively second person kind of thought, 'only third person and first person singular concepts, and concepts of those concepts' (p.245). We can characterise second personal linguistic communication as interpersonally self-conscious, the idea is, without positing a distinct layer of second person thought.

While Peacocke is fully explicit about the negative claim that there is no such thing as distinctively second person thought ('This description of what is involved in using and understanding the second person does not invoke a special second person concept or way of thinking' (p.245)), he says very little about how alternatively to construe the thoughts underlying sincere second person utterances. Assessment of the viability of holding together the views that second person utterances are interpersonally self-conscious and that there is no such thing as distinctive second person thought, then, calls for some ampliative interpretation. There are at least two things that Peacocke — or, as we might more safely call him, Peacocke* — might say.

The first is a position ascribed to Peacocke himself by M.G.F. Martin. This is the view that apparent uses of the second person concept in thought can be straightforwardly replaced with uses of *that person*. Any difference between the thoughts *you look nice* and *that person looks nice* must be located elsewhere than in a conceptual distinction between *you* and *that person*.

Martin develops an argument against this conceptual equivalence based on a difference in the behaviour of embedded first personal clauses in attitude reports with either *you* or *that person* featuring in the superordinate clause. There is a difference, for example, between the scope of the embedded first personal component in:

> (1) you think that I look nice

and in

(2) that person thinks that I look nice.[12]

In the first case, Martin argues, there are two ways of reading the scope of 'I'. It can either be read with what he calls a narrow scope, as constrained by the way in which the original thinker ('you') are thinking of me, or it can be read with a wide scope, as conveying nothing at all about the way in which you are thinking of me. In (2), by constrast, Martin discerns only a wide scope reading. There is no way of understanding the use of 'I' in the subordinate clause such that it conveys something about the way in which 'that person' is thinking of me.

The problem for a supporter of the conceptual equivalence of *you* and *that person*, according to Martin, lies in the attempt to explain this difference in scope between (1) and (2). The natural explanation of the unavailability of a narrow scope reading for (2) is that the audience has no access to a conceptual content associated with the embedded use of the first person that can be imputed to the thinker to which the thought is ascribed. We simply don't know how 'that person' is thinking of me; 'there is no concept associated with the speaker's current use of 'I' in attributing the thought to the [thinker] which the [thinker] could have been employing to think about the speaker's [looks].' (Martin 2014, p. 34) The explanation, by contrast, of the availability of a narrow-scope reading in (1) is that there *is* such access; 'some suitable concept is evoked in the speaker's use of 'I' which the audience may plausibly associate with the thinker, picked out by 'you', as employing in thinking the thought ascribed.' (Martin 2014, p. 34). In the case of (1), more specifically, it is quite natural to construe the content ascribed to the thinker as expressible with the sentence 'You look nice'.

So long as Peacocke* goes along with Martin this far, he is in trouble. For the thought expressible with the sentence 'You look nice', on Peacocke*'s view, is equally expressible using a third person demonstrative, 'That person looks nice' — for Martin's Peacocke*, there is no conceptual difference between these thoughts expressible in these two ways. The problem now, though, is that the thought expressed in the second way seems like a perfectly plausible candidate for content of the embedded thought in (2). This is a problem, because

[12]My example, not Martin's.

the explanation of the scope difference between (1) and (2) relied on the claim that there is no conceptual content associated with the embedded use of the first person in (2) that the audience could associate with the way the thinker is thinking of me. If there is no conceptual difference between the thought expressed with 'you' and 'that person', then it seems as if we no longer have a way of explaining why there are both wide- and narrow-scoped readings of (1), but only a wide-scoped reading of (2). It now seems as if we *do* have access to a concept plausibly associated with the thinker's ascribed thought in (2), which is equivalent to the concept associated with her thought in (1), so we are left without the means to say why the embedded first personal component behaves so differently in the two cases.

Whatever we make of Martin's argument, there is also something more general to say about this first way of construing Peacocke*'s positive view. It portrays him as accepting a fairly radical separation between second person language and the thoughts underlying them. If there is really no conceptual difference between formulating a given thought with *you* or *that person*, then the interpersonal self-consciousness indicated by a use of 'you' in language must be a purely linguistic phenomenon, with no links at all to the thought it expresses. There is no denying that this is a theoretical option; there are certainly other features of language that we take to be like this. At no point, however, does Peacocke himself deny any connection between the involvement of interpersonal self-consciousness in the use and understanding of second person language and thoughts it expresses — he says only that such interpersonal self-consciousness does not require us to posit a distinctive second person concept. Indeed, what he *does* say by way of positive characterisation of the thoughts underlying second person utterances seems to suggest a more moderate separation between thought and language. Rather than posit a distinctive second person concept, he says, we need 'only third person and first person singular concepts, and concepts of those concepts (and further concepts thereof, up the Fregean hierarchy)' (Peacocke 2014, ibid) Though somewhat programmatic, this remark is suggestive of a positive view that goes beyond a reduction of all apparent uses of *you* in thought to uses of third person demonstrative concepts.

What it suggests is a replacement of apparent uses of *you* in the formula-

tion of thoughts underlying second person utterances not with *that person*, but with the aggregate complex of concepts used in characterising the state of interpersonal self-consciousness expressed by second person utterances. This is Peacocke*'s second option. An utterer of the sentence 'you look nice', under this option, would be ascribed in full the thought that *the person I'm conscious of as being conscious of me being conscious of them looks nice*. This would leave Peacocke* with the materials to say that the interpersonal self-consciousness involved in the use and understanding of second person utterances also has reverberations at the level of thought.

The problem with this view, however, is that it positions us all rather improbably as philosophers of our own thoughts. Children are typically able to use the second person pronoun, in both possessive and non-possessive forms, by the age of three, and to comprehend its use by others even earlier.[13] This early use and understanding of second person language is tellingly juxtaposed with the cognitive demands of Peacocke*'s second view: it is difficult enough as theorists to keep track of the iterating layers of first and third person thought entering into the complex descriptive concept proposed on Peacocke*'s behalf as the way of thinking about others expressed by our sincere uses of the second person pronoun. Any view positing such capacities of ordinary thinkers — let alone toddlers — must surely lose any initial plausibility. Peacocke*'s second way of filling out the thoughts underlying our uses of 'you', then, maintains a connection between the interpersonal self-consciousness of second personal linguistic interactions with the thoughts underlying them only at the cost of glaring overintellectualisation.

A non-reductionist account of second person thought, by contrast, easily avoids this charge. This is because the relations of mutual awareness making up a state of interpersonal self-consciousness do not, on the non-reductionist picture, get into the thought's content. The only thought ascribed to the thinker is an irreducible *you*-thought — the thought, say, that *you look nice*. The awareness relations making up the state of interpersonal self-consciousness are no part of what the thinker thinks, but form, rather, part of the enabling conditions on second person thought. A thinker must *meet* (or at least take herself to

[13] See (Loveland 1984)

meet) the conditions specified by Peacocke in order to think a thought of this kind, but she need not have the capacity to conceptualise them.

Consider an analogy with perceptual demonstrative thought. To be in a position to think a *that*-thought about an object, a thinker must stand in a certain perceptual-attentional relation to it. This does not mean, however, that the thinker must represent that relation in the content of her thought — indeed, she need not even have the conceptual capacities to do so. The relation between the thinker and her object is merely an enabling condition on perceptual demonstrative thought, it is no part of the thought ascribed. Likewise, the idea is, for second person thought. The awareness relations constituting a state of interpersonal self-consciousness are no part of the thought ascribed, but stand as background conditions on thinking thoughts of that kind. Casting these relations in a mere enabling role, then, allows the non-reductionist to avoid the overintellectualisation charge facing Peacocke*'s second option; the non-reductionist attributes only a conceptually simple *you*-thought to the ordinary thinker.

By nevertheless bringing these relations into the picture in their enabling role, however, the non-reductionist retains the resources to say something about the role that this atomic concept *you* plays in our cognitive lives. Consider again an analogy with the perceptual demonstrative concept *that*. That a *that*-thought is canonically had in contexts of perceptual attention to an object shapes the normal patterns of use that we make of the concept. In chapter 1 we saw that feeling, smelling, hearing, tasting or seeing an object in a suitably attentive way all suffice to put one in a position to have a *that*-thought about it without drawing on any further information, or identificatory beliefs, about the object. That we use the concept in this way is hardly surprising in light of the enabling conditions on its use — given that perceptual attention is what *enables* my *that*-thought about the object, it is no wonder that so attending to the object puts me in a position to think about it in this way without drawing on any further information. The behaviour that is apt to immediately follow a deployment of *that* is likewise traceable, at least in part, to the enabling conditions on such thought. Having had a *that*-thought about an object I might reach out to touch it, or move my foot to avoid it, or any number of other actions depending on further facts about my standing motivational states. No matter how I

decide to react, however, the options made available by my *that*-thought are actions all that are rendered appropriate by the holding of a perceptual attentional relation to the object. The relation of perceptual attention need not be represented by the thinker of a *that*-thought, but even as a mere enabling condition on such thought its irrepressible influence over the normal patterns of use that we make of the concept is clear.

The same seems true of *you*. The relations involved in states of interpersonal self-consciousness need not, on the non-reductionist view, be any part of what is represented by the thinker of a *you*-thought, but as enabling conditions on such thought they plainly serve to fix, at least in part, the normal patterns of use that we make of the concept. It is only when a thinker is in (or, at least, *takes herself* to be in — more on this in §5.4.1 and §5.4.2) this network of conscious relations with another that she will be in a position to think about them immediately and without drawing on any further information about the situation, as *you*. That it is states of this kind feeding into our uses of *you* is only to be expected on the present account — given that the relation of interpersonal self-consciousness is what *enables* my *you*-thought about another, it is no surprise that my standing in such a relation positions me to think such a thought about them without drawing on any further identificatory beliefs about the object of my thought.

Similar observations can be made about the output side of the conceptual role for *you*, or the forms of behaviour rendered newly apt upon the deployment of a *you* concept. These will be many and varied, but share at their core an element of coordinated or joint action: we might move closer to one another to begin a conversation, or perhaps we will simply acknowledge the state of mutual recognition with a wave or a nod. This second aspect of the conceptual role for *you* is again perfectly intelligible in light of the enabling conditions on such thought on the present account. These are all actions of a kind that are made available by the state of interpersonal self-consciousness holding between the thinker and her referent. But given that such interpersonal self-consciousness is what *enables* the thinker's *you*-thought, it is only to be expected that such actions will become immediately available following such an episode of thought; their availability need not be mediated via an identificatory

belief. On the non-reductionist picture, then, the relations making up a state of interpersonal self-consciousness are no part of what is represented — or even represent*able* — by a user of the concept. Their status as background enabling conditions nevertheless provides the non-reductionist with valuable materials with which to account for the distinctive canonical patterns of use that we make of the second person concept.

The principal case for non-reductionism about *you*-thought issues from the twin desires to account for the interpersonal self-consciousness of second personal linguistic exchanges, and to avoid an overly intellectualised account of the thoughts underlying them. We are also, however, left with the side profit that this allows us to default to the intuition with which this section started: we can, on this view, rest easy with the idea that our second personal ways of talking track our second personal ways of thinking.

6.3 The argument from addressing (again)

Heck's challenge from §1.2 nevertheless remains. Second person thought, if it is to be understood as a proper coordinate of the second person pronoun, would seem to need to be characterised using the notion of addressing the other. Addressing, however, is an apparently purely linguistic phenomenon. Could there really be such thing as addressing someone in thought, for the non-telepathic creatures that we are? The question calls for more careful consideration of what is involved the act of addressing another. Until now we have been working with rough conception of it as a speech act in which the other notices that the utterance is directed towards them. I now want to suggest, however, that this working characterisation has been too conservative; what it takes to address someone is not as tightly bound to the use of language as this way of putting things suggests.

A natural way to proceed is to ask, what do we care about, when we care whether or not we have succeeded in addressing another? We care, or so it seems to me, whether or not the other is *receptive* to us; whether they are suitably sensitive to our attempts to engage their attention; whether we have made our attempts at connection sufficiently salient to them. In the standard case

of verbal communication we care whether they have noticed that one's utterance is directed towards them, but the phenomenon is just as familiar in cases of non-verbal interpersonal engagement. In non-verbal cases, it's just that we care whether we have succeeded in reaching out to them in other ways.

Jane Heal has recently gestured along similar lines that what is at issue, when we ask about addressing, is whether one's thought is adequately *open* to the other. She writes:

> It is not true [...] that one person's cognitive stances are all private, hidden from others, unless and until they get linguistic expression. Something may be common knowledge between agents, *where the thought of each is open to the other*, not in virtue of their speech but in virtue of their situation and/or their non-linguistic actions. And also it is not obviously true that, in such a situation of common knowledge, one agent cannot have a thought which is 'addressed' to the other, in some sense of 'address' appropriate to the thought having a second person character.

'So, for anything we have yet seen', she concludes, 'perhaps one person can 'address' another [...] without engaging in speech at all.' (Heal 2014, 320–321 (emphasis added)) At the core of the notion of addressing — or, at least, at the core of the notion of it bound up with discussions of the second person — is a kind of openness on the part of the agent, together with its recognition or receptivity on the part of the other. It is this that we care about in caring about in addressing another, and there is no reason to restrict it to recognisably linguistic exchanges.

Let me introduce a few examples to bring the point out. (1) My classmate and I have an ongoing private joke about a particular professor's habitual tardiness. We are in class together one day when, eventually, the professor in question turns up to his own lecture — twenty minutes late. We deliberately catch each others' eye to share a smile. (2) Waiting on the train platform, I sit side-by-side with a loved one thinking about our immanent separation. He presses my hand and I press his back. (3) My bedroom is immediately adjacent to my flatmate's. Between us we have devised a system of communication using wall-knocks to ascertain whether the other would like a cup of tea: three

knocks for an offer, two knocks for yes, one for no. I hear three knocks on the wall, and respond with a double tap. (4) My sister calls to me from across the room: 'do you want to come over tonight?'

It is, I suggest, initially plausible to think that there is addressing going on in each of these four cases. In only the last, however, is the interaction recognisably verbal. What this suggests, if it is right, is that the verbal form of addressing has received an undeserved level of attention in discussions of second person reference. We should think of it, rather, as just *one way* of carring out an act of addressing; a hand-squeeze or a wall tap will do just as well.

Such considerations bring along with them pressure to revise our understanding of addressing. I propose the following revision:

> *Addressing*: To address someone is to intend (to bring it about) that (i) they notice (or sustain notice) that one's attention is directed towards them and (ii) that they do so partly in virtue of recognising that very intention.

This is a formulation of two clauses. The first says that addressing someone is a act of attention, rather than of language. In addressing you, I try to get you to realise that I am attending to you. I can do this with words, but I can also do it with a hand-squeeze.

The second clause is in place to rule out counterexamples to the sufficiency of the formulation of the following kind. Imagine that you and I are at a conference dinner, seated only a few places away from a famous keynote speaker. In a shared attempt at flattery, we begin a loud discussion about the keynote's work while darting quick admirative glances in her direction. We are, here, intending that she notice that we are attending to her, but it does not yet seem like a case of addressing. To borrow an expression from the Gricean communicative intention literature, for that, there must be no 'sneaky intentions' of the kind built into this case: to address her, we must intend that the keynote recognise that our attention is directed towards her, but we must also intend that she does so partly by a recognition of that very addressive intention.

Under this revised characterisation, there is no obstacle to the verdict that cases (1)-(3) involve somebody addressing someone else, just as much as (4). It also removes any obstacle to the claim that second person thought refers, just

like its linguistic counterpart, to its addressee. With this liberalised characterisation of addressing in place, this is just to say that it refers to the person one is both attending to, and intending that they notice that one is attending to them (in virtue of recognising that very intention). No telepathy required.[14]

Indeed, given its characterisation in the last section as interpersonally self-conscious thought, it is really no surprise that second person thought refers to its addressee, understood in the revised way. Another way of formulating the proposed new characterisation of addressing, after all, would be to say that it is an act of attempting to induce a state of interpersonal self-consciousness with the other person — it is the attempt to get them to notice one's occurrent awareness of them. To say that second person thought refers to its addressee, then, is just to say that *interpersonally self-conscious thought is thought about the person with respect to whom one has attempted to induce a state of interpersonal self-consciousness.*

At the end of the last section, we saw that the non-reductionist has the resources to fill in and account for some of the central aspects of the conceptual role for the second person concept. Space has now been created in the current section to also give a non-reductive account of the rule of reference determination for that concept too. So long as we are happy to relax the conditions on the notion of addressing in the way rendered plausible by cases (1)-(4), we can offer the rule that uses of *you* refer to their addressees. The argument from addressing no longer blocks the way to a non-reductive view of second person thought.

6.4 Some counterexamples

There are at least three kinds of possible counterexample the view just outlined.

[14]The issue of whether there is addressing going on in these cases should not be mistaken for a verbal dispute. The important claim is that there is a certain sort of mind-to-mind interaction in each of the pairs involved in these cases (and others like them) of a kind that makes second person thought available. If this strikes one as stretching the term 'addressing' too far, then an alternative framing of the same point would be to say that (verbal) addressing is merely one version of the interpersonal stance characteristic of second person interactions; these cases demonstrate that there are others.

6.4.1 Error cases

The first are cases in which I take myself to have induced a state of interpersonal self-consciousness with respect to another, but am mistaken — my classmate in (1), for instance, might be aiming her smile at another friend behind me, so there is, in fact, no state of interpersonal self-consciousness holding between us. Suppose now that I essay a second person thought about her. Are we forced to the view that I am not really thinking a second person thought here at all?

At first pass, this needn't be an unwanted result. In other areas of philosophy of mind we have become familiar with the idea that the availability of certain kinds of thought depends on the cooperation of worldly facts — on some views of singular thought, whether or not an object exists, for instance, will affect whether or not it will be possible to think a genuine object-dependent singular thought about it.[15] It might seem to be no great stretch to say that the possibility of thinking a genuine *you*-thought likewise depends on the cooperation of worldly facts, only that in this case it is a special kind of fact, viz., facts about the mental activities of another. Just as the existence or non-existence of an object can determine the availability of object-dependent thought, the idea would be, the participation or non-participation of one's referent in a state of interpersonal self-consciousness can affect the availability of second person thought. It seems right, moreover, that there is *something* defective about these cases; biting the bullet in this way is one way to accommodate this intuition.

There is, however, an important respect in which the point about object-dependent singular thought differs from the one suggested in the case of *you*-thought. That is, we might think that facts about whether or not one's object of thought exists pack a weightier punch with regards to the thoughts one can have about it than mere facts about how things are with it. Pressure in the case of object-dependent singular thought to say that where there is no existing referent there is no thought of this kind at all can be seen to come from the idea that there is, in such cases, no way of specifying *what it would be* for the thought to be true. This means that any truth-conditional theory of thought content

[15] If such thoughts there are. I do not presume here to settle questions about whether singular thoughts are really object-dependent, or if so, how empty singular thoughts should be dealt with; I only draw on some of the ideas made familiar by these discussions.

will run into difficulties in saying how we should understand the thought's content. Things are very different with the suggestion that the availability of second person thought depends on whether or not the referent is engaged in a state of mutual awareness with the thinker. We would still know *what it would be* for my thought of the form *you are F* to be true. Whatever it is that is defective about my thought, it does not come down to the fact that it has no specifiable truth conditions. There is no corresponding pressure, then, to concede that a difference in how things are mentally with one's object of thought will affect whether or not one can have a genuine second person thought about them.

Another look at the revised characterisation of addressing brings out a different way of responding to these cases. We need only notice that under the proposed revision, addressing does not depend on one's *success* in drawing the other's attention: one can address someone by merely *intending* that they notice that one is attending to them. One might well fail in that intention, but nevertheless still be counted as addressing them. The described case above seems to be one of just this kind. I am intending that my friend notice that I am attending to her, and, indeed, (erroneously) take myself to have succeeded in carrying out this intention.

My failure leaves the success of my thought intact. We can now say that I have a second person thought about my friend that refers to its addressee, where its addressee is the person with respect to whom I have intended to induce a state of interpersonal self-consciousness. My failure, however, also accounts for the sense that I have gone wrong in some way. I have failed to pull off the intention at the very heart of the kind of thought I am having and, what's more, I am ignorant of my failing. The thoughts involved in cases of this kind might be flawed second person thoughts, but they are second person thoughts nonetheless.

6.4.2 'As if' cases

A second kind of counterexample is much more deliberate on the part of the thinker. I might knowingly and intentionally seem to address someone (or something) in a circumstance in which a state of interpersonal self-

consciousness is manifestly impossible. This is familiar phenomenon. I might think *I'm so proud of you* as I behold my sleeping child, or mentally execute a triumphant cry of *there you are!* upon finding my keys. Plausibly, this is something we do all the time for comic effect, or as a way of testing one's emotional stance about something, or to stave off loneliness, or to give vent to otherwise pent up feelings. Clearly, there is no attempt to induce a state of interpersonal self-consciousness here and so, on the present account, no second person thought.

One way to go on these cases would be to give them a similar treatment to the earlier example, in which I internally articulated a second person utterance to a fellow shopper. They should be treated, the idea would be, as cases of inner speech, in which I act *as if* I was engaging in a second personal linguistic exchange with the other. This is an approach articulated by M.G.F. Martin, who writes:

> In ordinary linguistic communication a certain psychological structure is present: speaker and audience are related in terms of mutual awareness. [...]. In other situations, [...] the thinker treats their actual situation as if it were like that present in a core case of linguistic communication. (Martin 2014, p. 33)

Should we count these as episodes of second person thought? The question is not straightforward to adjudicate. On the one hand, if the target phenomenon is a non-reducible kind of thought whose canonical linguistic expression takes the form of second person pronouns, then it seems quite right to say that what we have here are second person thoughts. The central involvement in these cases of the faculty of imagination, on the other hand, might tempt us to a treatment of them as something less than sincere or genuine *you*-thoughts. It may simply turn out that once the relevant phenomena have been properly characterised, the question of where we draw the line will be a matter of taste.

We have, in fact, already confronted a question of this kind in §5.2 in the context of the discussion of the case of Hermione and Leontes. Hermione's mere imagining of the obtaining of the fourth condition involved in a state of interpersonal self-consciousness, we saw, was enough to release her capacity

to think thoughts of a kind that she could not think before. It seems to me that cases like this make it very natural to think that an imagined enactment of a state of interpersonal self-consciousness with another is enough to provide a thinker with the right pscyhological structures within which genuine second person thought becomes available. The cases described here are, as we might put it, the more deliberate cousins of the error cases of §5.4.1 — second person thoughts all, even if both these and the error cases are also in some sense derivative or secondary to a case in which the conditions on a state of interpersonal self-consciousness really are all met.

Another sort of reaction to cases of this kind is given by Peacocke. Upon seeing the erratic behaviour of another driver, he imagines entertaining the thought 'If you go on driving like that, you will be involved in an accident' (Peacocke 2014, p. 248). This thought, according to Peacocke, loses nothing by a reformulation as a *that-person* thought, or as a thought had under a perceptually based mode of presentation. Is this a better explanation of the phenomenon at hand? There is, I think, much to be said for a liberal stance here: we are perfectly free to allow that there can be cases and cases.

There is at least one *prima facie* advantage, however, to making room for the enactment response in at least some of these cases. That is, that it has something to say about why it is that there is something *theatrical* about thoughts of this kind. Under the proposed view, that is because there really is a kind of play-acting going on here; the thinker is, for whatever reason, occupied in an imaginative project of acting out a state of interpersonal self-consciousness with respect to someone (or something) in which second person thought about them (or it) would be available. Peacocke's proposal, by contrast, cannot fall back on the same resources to explain this feature of these thoughts. If they really are just perceptual demonstrative thoughts differently expressed, then it is not clear how we should explain the sense that there is something staged, or acted, about thought episodes of this kind.

6.4.3 Calling someone's attention

The third kind of case is one in which I give voice to a second person utterance in order to attract someone's attention with the phrase 'hey you!'. Given the attempt to attract the other's attention here, the use of the second person pronoun cannot be expressive of a prior state of interpersonal self-consciousness. It cannot be the case, then, that I take myself to have met the enabling conditions on second person thought.

On the proposed account of this chapter, it is clear that this use of the second person pronoun cannot be expressive of an underlying second person thought. It seems equally clear, however, that this verdict does nothing to undermine that account. Pretheoretically, it seems *right* to say that this use of the second person pronoun is not expressive of a second person thought, or indeed of any thought at all — rather, it is quite natural to think that the word is being used merely for its acoustic properties. The speaker is using the sound of the utterance to attract the attention of the hearer, not to express an underlying thought.

If this really is a plausible reading of the case, then there may even be a way to trade in its status as counterexample for one of support for the present view. If the word is being used merely for its acoustic properties, then any word (or, indeed, sound) would do. Why, then, the convention of reaching for a second person pronoun? One explanation is that in attracting the other's attention I am attempting to induce the very state manifested in a fully sucessful second person thought. The use of the pronoun 'you' is anticipatory; it is used in expectation of the thought that will become available once the state of interpersonal self-consciousness has been established. Far from undermining the present account, then, there is a natural reading of this case under which it seems to add to its credibility.

6.5 From you to me again

This chapter began with an observation from McDowell about what it takes to understand second personal utterances. His example showed that in order to understand a second person utterance directed towards me, I must think

a first person thought. In §5.1.2 we saw that together with a naïve view of communication, these understanding constraints could be accommodated by a certain form of reductionism about second person thought — namely, the view that second person thought is identical to first person thought. If the kind of understanding required for communication involves the sharing of a single thought between speaker and hearer, and if first and second person thought are really the same thought, then it is only to be expected that understanding a second person utterance will require uptake via a first person thought. In that section it was flagged that a rejection of this explanation of the McDowellian understanding constraints gives rise to an explanatory debt. With our non-reductionist picture of second person thought in place, do we now have the resources to redeem that debt?

I think we do. Just such an explanation becomes available by the characterisation of second person thought as thought about its addressee, where the notion of addressing is given the treatment of §5.3. That is, it becomes available on a view on which second person thought is thought that is about the person with respect to whom the thinker has attempted to induce a state of interpersonal self-consciousness. Earlier, I responded to Michael Thompson's objection that no amount of stacking of neutrally formulated 'doxastic pancakes' could suffice to characterise a state of interpersonal self-consciousness. It's not enough to say that there is mutual awareness between thinker and referent, and mutual awareness of that mutual awareness. I emphasised that to capture the special intimacy of a state of interpersonal self-consciousness, it is crucial that these states of awareness are filled in in a certain kind of way. In attempting to induce a state of interpersonal self-consciousness, I, as thinker, attempt to bring it about that you are aware of me (thought of *as me, myself*) being aware of you (thought of *as you, yourself*). On both ends of a successful act of addressing, we must each think of ourselves *as ourselves*.

What this means is that it is built into the very conditions on addressing another person that the other is self-aware in a distinctively first personal way. There is a sense, then, in which the conditions on second person thought subsume the conditions on first person thought on the part of the referent. McDowell's understanding constraints are thus explicable by the fact that part of

what it is to have a second person thought, when all goes well, is for the referent to have a self-conscious thought about themselves. To articulate these understanding constraints on second person utterances is just to make explicit part of the conditions underlying the thought expressed, only from the unique perspective of the addressee. When you, as addressee, understand the thought that has been expressed, the way to articulate what was involved in that thought from your side is that *you are aware of me, myself being aware of you, yourself (and are thinking that I am F).* No wonder, then, that understanding my utterance requires thinking a first person thought on your part.

Explanation of McDowell's understanding constraints is a desideratum on any account of the thoughts underlying second person utterances. In §6.1.1 we saw that a particular reductionist view supplied such an explanation by identifying first and second person thought. It now seems, however, that there is a way of explaining them on a non-reductionist view too. The interconnectedness between first and second person thought with which we started is just as naturally explained by a view on which there is such thing as distinctive second person thought.

6.6 Conclusion

The aim of this chapter was to elaborate and defend a non-reductive account of second person thought, and to explain its interconnectedness with first person thought. The emerging picture has been one on which second person utterances are expressive of distinctive second person thoughts, which refer, just like their linguistic counterparts, to their addressees. One of the central proposals of this chapter has been that we need to relax our understanding of what it takes to address someone, that it should not be understood as a purely linguistic phenomenon. It should be understood, rather, as the act of attending to someone while intending that they notice that one's attention is directed towards them (by recognition of that very addressive intention), or as the act of attempting to induce a state of interpersonal self-consciousness with respect to the other person. This losening of the notion of addressing allows us to say that second person thought is interpersonally self-conscious thought that refers to

the person with respect to whom one has attempted to induce a state of interpersonally self-conscious thought.

Second person thought is distinctive because it is distinctively interpersonally self-conscious thought: just as no third person concept could capture the self-consciousness of first person thought, no third or first person concept (or combination of the two) could serve to capture the interpersonal self-consciousness of second person thought.

Chapter 7

Conclusion

The machinery of this dissertation is not new. Most of the moving parts of the account of first person thought developed over the first five chapters are shared with many of the authors whose influence is felt in the final form of that account, or whose work I have used as foils in bringing out its contours. Some of this common ground might have sometimes been lost where differences in our ways of systematising the shared components issue in different-looking resulting theories. Even where they have not, it will be illuminating at this end point to bring some of those points of agreement and disagreement back into the foreground.

The foundational idea of this dissertation, that we must look to our most basic forms of self-awareness in order to explain our capacity for first person thought, is inherited from O'Brien. With her I share the conviction that the topics of self-reference of self-knowledge cannot be easily separated. So long as our target is first person thought itself rather than a psychologically idealised etiolated abstraction from it, then we must situate our account of how we think of ourselves reflexively in a wider context of what we are like as conscious, and self-conscious, creatures. Where we disagree is on what these basic forms of self-awareness will look like. For O'Brien, the form of awareness playing this key role cannot be anything less than our distinctive awareness or our own intentional actions. On my view, there will be as many different forms of awareness playing this role as there are differently built creatures with different forms of essentially subject-reflexive ways of knowing about themselves. First person

thought need not, in principle, be out of the reach of a non-agentive thinker, if such a thinker there could be.

With Campbell I share some areas of solid ground. We are agreed that the self reference rule (his token reflexive rule) is not to be touched, and we are also agreed that there are distinctively first personal forms of awareness that need housing in a full account of first person thought. We merely part ways on how this is to be done. For Campbell, there is no justifying, either causally or normatively, the fact that we use the first person in immediate response to some, but not all, of our ways of knowing about ourselves — or at least, there is no justifying it by appeal to the relevant rule of reference. In the face of the challenge of chapter 1 we must simply throw up our hands and rest content with a picture on which the patterns of use that we make of the first person concept are treated as primitive. My view is more optimistic. We *can* explain our patterns of use by the self reference rule; we must merely take a bypass through the epistemic conditions on first person thought engendered by that rule.

The points of communication with Peacocke's views have been more visible than some of the rest, particularly in setting up the starting explanatory challenge in chapter 1. We are in agreement that the first person concept's conceptual role is ultimately grounded in the fundamental condition for something to be the referent of a use of that concept. But in the end, my account has our first personal forms of awareness play a special role in the full story of first person thought that they do not play on his. This, I think, provides me with new resources to answer our shared starting challenge that he must do without. The writings of Campbell and Peacocke have both been instrumental in shaping my understanding of the notion of immunity to error through misidentification, and the kinds of explanations it is subject to, as set out in chapter 3. Peacocke is also, of course, my principal opponent in the chapter on second person thought.

Given his use (and my non-use) of a mental files framework, it might be a little harder to make out the overlap with Recanati's account of first person thought. There is an important sense, however, in which we are trying to do something very similar. For both of us, the way to answer questions about self-

reference, self-knowledge and their interconnectedness, is by appeal to something more basic than either an epistemic or a referential relation. More important than either of these is the relation of identity. For Recanati, this idea takes shape in the claim that identity is the epistemically-rewarding base relation that individuates the 'self' file, a relation that both determines the reference of that file and that determines which particular epistemic relations will serve to provide immediate information about that referent. On my view, identity is the overarching constraint on both the referential and the epistemic conditions on fully comprehending first person thought. For us both it is explanatorily more fundamental than any of the particular contingent forms of self-awareness that happen to correspond in the right way to this identity relation for the creatures that we are. This elevation of the identity relation to centre stage is also a feature of the account we share with Rödl. For Rödl, it is only in virtue of our bearing this relation to ourselves that we are in a position to have spontaneous, rather than merely receptive knowledge of ourselves, and it is only by the possession of spontaneous self-knowledge that we can attain the logical perspective on ourselves afforded by self-reference. For Rödl, then, like for Recanati and like for me, at the very bottom of an account of self-knowledge or self-reference must be the bare metaphysical relation of identity.

A second correspondence between my account and Recanati's is also worth noting. In chapter 2, I offered a characterisation of a cognitive achievement that I thought was typically present in our singular thoughts. I described this achievement as one of grasping one's own use of a concept, or of undergoing a fully comprehending episode of thought. To achieve it, I said, one must be in a position to know empirical facts of the form *a* is *F* about it, under *a*. This seems to tally, on Recanati's account (and other mental file theorists before him) with what it is to employ a mental file to think about an object. Files contain predicates. These predicates record information concerning the object that the file is about. So even to open a file on an object, one must be in a position to predicate something of it. The full story on Recanati's account of what it is to deploy the 'self' file, then, largely coincides in this respect with the full story on mine of what it is to think a first person thought with full comprehension.

I have left until last the most important influence and foil of the account. In

chapter 1 I dismissed with little ceremony the so-called demonstrative model of first person thought, held in its canonical form by Evans. Problems with that account, I said, are well known. A fair framing of the undertaking of this dissertation, however, is as an attempt to make up for the costs of that rejection. By allowing the determination of first person reference to be settled by our special forms of self-awareness, the demonstrative model theorist is rich with resources with which to easily explain such phenomena as the canonical patterns of use that we make of the first person concept, or the asymmetry of first personal immunity to error through misidentification. In casting aside that model we also cast aside those resources. The project of this dissertation has been to see whether, by repositioning some of these moving parts, we can reinstate some of the benefits of a demonstrative model of first person thought without sacrificing the self reference rule model of reference determination.

Bibliography

Anscombe, Elizabeth (1997). "The First Person". In: *Self-Knowledge*. Ed. by Q. Cassam. Oxford University Press. Chap. The first person.

Armstrong, D.M. (1984). *Consciousness and Causality*. Ed. by D.M. Armstrong and N. Malcom. Basil Blackwell.

Ayers, Michael (1991). *Locke: Epistemology and Ontology*. Routledge.

Bartlett, F.C. (1932). *Remembering: A study in Experimental and Social Psychology*. Cambridge University Press.

Bell, David Andrew (1979). *Frege's Theory of Judgement*. Oxford University Press.

Bermudez, J. (2005). "Evans and the Sense of "I"". In: *Thought, Reference, and Experience: Themes from the Philosophy of Gareth Evans*. Ed. by J. Bermudez. Oxford University Press. Chap. Evans and the Sense of "I".

Bermudez, J.L. (2013). "Immunity to error through misidentification and past-tense memory judgements". In: *Analysis* 73.2, pp. 211–220.

Bigna Lenggenhager Tej Tadi, Thomas Metzinger Olaf Blanke (2007). "Video Ergo Sum: Manipulating Bodily Self-Consciousness". In: *Science* 317.5841.

Botvinick, Matthew and Jonathan Cohen (1998). "Rubber hands'feel' touch that eyes see". In: *Nature* 391.6669, pp. 756–756.

Brewer, B. (1995). "Bodily Awareness and the Self". In: *The Body and the Self*. Ed. by J.L. et al. Bermudez. MIT Press. Chap. Bodily Awareness and the Self.

Burge, T. (1979). "Sinning against Frege". In: *The Philosophical Review* 88, pp. 398–432.

Campbell, John (1994). *Past, Space and Self*. MIT Press.

— (1999a). "Immunity to Error through Misidentification and the Meaning of a Referring Term". In: *Philosophical Topics* 26, pp. 89–104.

Campbell, John (1999b). "Schizophrenia, the space of reasons and thinking as a motor process". In: *The Monist*.

— (2012). "On the thesis that 'I' is not a referring term". In: *Immunity to error through misidentification: New Essays*. Ed. by S. Prosser and F. Recanati. Cambridge University Press.

Cassam, Quassim (1997). *Self and World*. Oxford University Press.

Chen, Cheryl (2011). "Bodily awareness and immunity to error through misidentification". In: *European Journal of Philosophy*.

Claudia Lunghi Maria Concetta Morrone, David Alais (2014). "Auditory and Tactile Signals Combine to Influence Vision during Binocular Rivalry". In: *The Journal of Neuroscience* 34.3, pp. 784–92.

Coliva, Annalisa (2002). "Thought insertion and immunity to error through misidentification". In: *Philosophy, Psychiatry, and Psychology* 9.1, pp. 27–34.

— (2006). "Error through Misidentification: Some Varieties". In: *The Journal of Philosophy* 103.8, pp. 403–425.

— (2012). "Which 'key to all mythologies' about the self? A note on where the illusions of transcendence come from and how to resist them". In: *Immunity to error through misidentification: New essays*. Ed. by Simon Prosser and Francois Recanati. Cambridge University Press, pp. 22–45.

Davies, Martin (1982). "Individuation and the Semantics of Demonstratives". In: *Journal of Philosophical Logic* 11.3, pp. 287–310.

De Vignemont, Frédérique (2011). "A Self for the Body". In: *Metaphilosophy* 42.3, pp. 230–247.

— (2012). "Bodily Immunity to Error". In: *Immunity to Error through Misidentification*. Ed. by F. Recanati and S. Prosser. Cambridge University Press.

— (2013). "The Mark of Bodily Ownership". In: *Analysis* 73.4, pp. 643–651.

Dickie, Imogen (2011a). "How Proper Names Refer". In: *Proceedings of the Aristotelian Society* 111.1pt1, pp. 43–78.

— (2011b). "Visual Attention Fixes Demonstrative Reference by Eliminating Referential Luck". In: *Attention: Philosophical and Psychological Essays*. Ed. by Declan Smithies Christopher Mole and Wayne Wu. Oxford University Press.

Dickie, Imogen and Gurpreet Rattan (2010). "Sense, Communication, and Rational Engagement". In: *Dialectica* 64.2, pp. 131–151.

Egan, Andy (2007). "Epistemic Modals, Relativism and Assertion". In: *Philosophical Studies* 133.

Erhsson H.H., Holmes N.P. Passingham R.E. (2005). "Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas". In: *Journal of Neuroscience* 25.45.

Evans, Gareth (1981). "Understanding Demonstratives". In: *Meaning and Understanding*. Ed. by Herman Parret. Clarendon Press, pp. 280–304.

— (1982). *The Varieties of Reference*. Ed. by J. McDowell. Oxford University Press.

— (1985). "Molyneux's Question". In: *Collected Papers*. Ed. by Gareth Evans. Oxford University Press.

Fernández, Jordi (2014). "Memory and Immunity to Error Through Misidentification". In: *Review of Philosophy and Psychology* 5.3, pp. 373–390.

Gallagher, Shaun (2000). "Self-reference and schizophrenia: a cognitive model of immunity to error through misidentification". In: *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-Experience*. Ed. by Dan Zahavi. Vol. 1. John Benjamins, pp. 203–239.

GF Misceo WA Hershberger, RL Mancini (1999). "Haptic estimates of discordant visual-haptic size vary developmentally". In: *Percept Psychophys.* 61.608-14.

Hamilton, Andy (2009). "Memory and self-consciousness: Immunity to error through misidentification". In: *Synthese* 171.3, pp. 409–417.

Hawthorne, John and David Manley (2012). *The Reference Book*. Oxford University Press.

Heal, Jane (2014). "Second Person Thought". In: *Philosophical Explorations* 17.3, pp. 317–331.

Heck, Richard (2002). "Do Demonstratives Have Senses?" In: *Philosophers' Imprint* 2.2, pp. 1–33.

Howell, R.J. (2011). "Immunity to error through misidentification and subjectivity". In: *Canadian Journal of Philosophy* 37.4, pp. 581–604.

Kriegel, Uriah (2009a). "Self-representationalism and phenomenology". In: *Philosophical Studies* 143.3, pp. 357–381.

Kriegel, Uriah (2009b). *Subjective Consciousness: A self-representational theory*. Oxford University Press.

Langland-Hassan, Peter (2015). "Introspective Misidentification". In: *Philosophical Studies* 172.7, pp. 1737–58.

Longworth, Guy (2013). "Sharing Thoughts about Oneself". In: *Proceedings of the Aristotelian Society*.

— (2014). "You and Me". In: *Philosophical Explorations* 17.3, pp. 289–303.

Loveland, Katherine A. (1984). "Learning about points of view: spatial perspective and the acquisition of I/you". In: *Journal of Child Language* 11.3, pp. 535–556.

MA Heller JA Calcaterra, SL Green L Brown (1999). "Intersensory conflict between vision and touch: the response modality dominates when precise, attention-riveting judgments are required". In: *Percept Psychophys.* 61.7, pp. 1384–98.

Mach, Ernst (1914). *The Analysis of Sensations and the Relation of the Physical to the Psychical*. Ed. by s. Waterlow. Trans. by C.M. Williams.

Macpherson, Fiona (2011). "Cross-Modal Experiences". In: *Proceedings of the Aristotelian Society* 111.3pt3, pp. 429–468.

Maravita, A. (2008). "Spatial Disorders". In: *Cognitive Neurology: a clinical textbook*. Ed. by J.F. Demonet P.C. Fletcher S.F. Cappa J. Abutalebi. Oxford University Press.

Martin, C. B. and Max Deutscher (1966). "Remembering". In: *Philosophical Review* 75.April, pp. 161–96.

Martin, M. G. F. (1992). "Perception, Concepts and Memory". In: *The Philosophical Review* 101.4, pp. 745–763.

— (1995). "Bodily awareness: a sense of ownership". In: *The Body and the Self*. Ed. by J. et al Bermudez. MIT Press.

— (2014). "In the Eye of Another: Comments on Christopher Peacocke's 'Interpersonal Self-Consciousness'". In: *Philosophical Studies* 170.1, pp. 25–38.

McDowell, John (1984). "*De re* senses". In: *The Philosophical Quarterly* 34.136, pp. 283–294.

— (1994). "Mind and World". In: chap. Chapter 3.

Merleau-Ponty, M. (1989). *Phenomenology of Perception*. Routledge.

Morgan, Daniel (2009). "Can You Think My 'I'-Thoughts?" In: *Philosophical Quarterly* 59.234, pp. 68–85.

— (2015a). "The Demonstrative Model of First-Person Thought". In: *Philosophical Studies* 172.7, pp. 1795–1811.

— (2015b). "Thinking About the Body as Subject".

Nadia Bolognini Carlo Cacchetto, Carlo Geraci Angelo Maravita Alvaro Pascual-Leone Costanza Papagno (2012). "Hearing shapes our perception of time: temporal discimination of tactile stimuli in deaf people". In: *Journal of Cognitive Neuroscience* 24.2, pp. 276–286.

Ninan, Dillip (2010). "De se attitudes: ascription and communication". In: *Philosophy Compass* 4, pp. 1–16.

O'Brien, Lucy F. (1995). "Evans on Self-Identification". In: *Noûs* 29.2, pp. 232–247.

— (2007). *Self-Knowing Agents*. Oxford University Press.

— (2011). "Consciousness and the Self". In: *Ordinary Self Consciousness*. Ed. by John Perry and JeeLoo Liu. Cambridge University Press, pp. 101–122.

— (2012). "Action and immunity to error through misidentification". In: *Immunity to Error through Misidentification: New essays*. Ed. by François Recanati and Simon Prosser. Cambridge University Press.

Peacocke, Christopher (1981). "Demonstrative Thought and Psychological Explanation". In: *Synthese* 49.2, pp. 187–217.

— (1992). *A Study of Concepts*. MIT Press.

— (1998). "Philosophy of Language". In: *Philosophy 2: Further Through the Subject*. Ed. by A. C. Grayling. Oxford University Press.

— (2008). *Truly Understood*. Oxford University Press.

— (2012). "Explaining de se phenomena". In: *Immunity to Error through Misidentification: New essays*. Ed. by Simon Prosser and Francois Recanati. Cambridge University Press, pp. 144–157.

— (2014). *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness*. Oxford University Press.

Perry, John (1977). "Frege on Demonstratives". In: *The Philosophical Review* 86.4, pp. 474–497.

— (1990). "Self-Notions". In: *Logos*, pp. 17–31.

Prosser, Simon (2012). "Sources of Immunity to Error through Misidentification". In: *Immunity to Error through Misidentification: New Essays*. Ed. by Simon Prosser and Francois Recanati. Cambridge University Press, pp. 158–179.

Pryor, J. (1999). "Immunity to error through misidentification". In: *Philosophical Topics* 26, pp. 271–304.

Ramachandran, V. S. and William Hirstein (1998). "The perception of phantom limbs: the D.O. Hebb lecture". In: *Brain* 121, pp. 1603–1630.

Recanati, François (1993). *Direct Reference: From Language to Thought*. 178. Blackwell, p. 134.

— (2012a). "Immunity to error through misidentification: what it is and where it comes from". In: *Immunity to Error through Misidentification: New Essays*. Ed. by Simon Prosser and Francois Recanati. Cambridge University Press, pp. 180–201.

— (2012b). *Mental Files*. Oxford University Press.

Reichenbach, Hans (1947). *Elements of Symbolic Logic*. The MacMillan Company.

Roache, Rebecca (2006). "A defence of quasi-memory". In: *Philosophy* 2.323-355.

Rödl, Sebastian (2007). *Self-Consciousness*. Harvard University Press.

Roedinger, H.L. and K.A. DeSoto (2015). "The psychology of reconstructive memory". In: ed. by J. Wright. 2nd ed. Elsevier. Chap. International encyclopedia of the social and behavioural sciences.

Roland Zahn, Jochen Talazko and Dieter Ebert (2008). "Loss of the Sense of Self-Ownership for Perceptions of Objects in a Case of Right Inferior Temporal, Parieto-Occipital and Precentral Hypometabolism". In: *Psychopathology* 41, pp. 397–402.

Sainsbury, R.M. and M. Tye (2012). *Seven Puzzles of Thought and How to Solve Them: an originalist theory of concepts*. Oxford University Press.

Schachter, Daniel (1999). "The Seven Sins of Memory". In: *American Psychologist* 54.3.

Schear, Joseph K. (2009). "Experience and Self-Consciousness". In: *Philosophical Studies* 144.1, pp. 95–105.

Schechtman, Marya (1996). *Constitution of Selves*. Cornell University Press.

— (2010). "Memory and Identity". In: *Philosophical Studies* 153.1, pp. 65–79.

Schwitzgebel, Eric (2014). "Introspection". In: *The Stanford Encyclopedia of Philosophy*.

Shimada, Hiroyuki (1990). "Effect of auditory presentation of words on color naming: the intermodal Stroop effect". In: *Perceptual and Motor Skills* 70, pp. 1155–61.

Shoemaker, Sydney (1968). "Self-Reference and Self-Awareness". In: *Journal of Philosophy* 65.19, pp. 555–567.

— (1970). "Persons and Their Pasts". In: *American Philosophical Quarterly* 7.4, pp. 269–85.

Smith, A. D. (2002). *The Problem of Perception*. Oxford University Press.

Smith, Joel (2006). "Bodily Awareness, Imagination and the Self". In: *European Journal of Philosophy*.

Soteriou, Matthew (2013). *The Mind's Construction: the ontology of mind and mental action*. Oxford University Press.

Strawson, P. F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. 2. Routledge, pp. 246–246.

Sutton, John (2011). "Influences on memory". In: *Memory Studies* 4.4, pp. 355–359.

Sutton, John and Carl Windhorst (2009). "Extended and Constructive Remembering: two notes on Martin and Deutscher". In: *Crossroads* IV.1.

T Seizova-Cajic, R Azzi (2011). "Conflict with vision diminishes proprioceptive adaptation to muscle vibration". In: *Experimental Brain Research* 211.2, pp. 169–175.

T Seizova-Cajic, WL Sachtler (2007). "Adaptation of a bimodal integration stage: visual input needed during neck muscle aftereffect". In: *Experimental Brain Research* 181.117-29.

Thompson, Michael (2012). *You and I*. audio podcast, Aristotelian Society.

V Occelli C Spence, M Zampini (2013). "Auditory, tactial and audiotactile information processing following visual deprivation". In: *Psychological Bulletin* 139.1, pp. 189–212.

Wittgenstein, Ludwig (1958). *The Blue and Brown Books: Preliminary studies for the 'Philosophical Investigations'*. Blackwell Publishing.

Wright, Crispin (2012). "Reflections on Francois Recanati's 'Immunity to Error through Misidentification: what it is and where it comes from'". In: *Immunity to Error through Misidentification: New essays*. Ed. by Simon Prosser and Francois Recanati. Cambridge University Press, pp. 247–280.

Y Jiang, L Chen (2013). "Mutual influences of intermodal visual/tactile apparent motion and auditory motion with uncrossed and crossed arms". In: *Multisensory Research* 26.1-2, pp. 19–51.

Zahavi, Dan (2005). *Subjectivity and Selfhood: investigating the first-person perspective*. Cambridge MA: MIT press.