BEING

ORIGINS

FLORIAN

WÜST

HOLZ

FLORIAN L. WÜSTHOLZ

BEING ORIGINS

BEING ORIGINS

The way we think about ourselves

FLORIAN L. WÜSTHOLZ

Dissertation zur Erlangung der Doktorwürde
an der Philosophischen Fakultät der Universität Freiburg, CH.

Genehmigt von der Philosophischen Fakultät auf Antrag
des Herrn Professors Gianfranco Soldati (1. Gutachter)
und der Frau Professorin Lucy O'Brien (2. Gutachterin).

Freiburg, den 14. Oktober 2017, Prof. Bernadette Charlier, Dekanin.

Florian L. Wüstholz: *Being Origins*
Fribourg, April 2018

This thesis is licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International License.

This means that you're free to share this thesis by copying and redistributing it in any medium or format as long as you give appropriate credit to the author, don't use the material for commercial purposes and don't remix, or transform it.

Dedicated to Pascale Anna.

ACKNOWLEDGMENTS

Many people will probably be familiar with Karl Marx's 11th thesis on Feuerbach in some way or other. If you're not among them, worry not: 'The philosophers have only *interpreted* the world in various ways; the point is to *change* it' (Marx and Engels 1975: 5). How can this dissertation change the world? I don't know and I doubt it can. With war still raging in Syria, Yemen, Iraq, Afghanistan, Somalia, Darfur and several other areas of the world, with countries being ruled by dictators, almost-dictators, demagogues, and pathological narcissists, with democracies being undermined and attacked by right-wing totalitarian, fascistic, nationalistic, and populist movements, with more than 150'000'000'000 (i.e. 150 billion) nonhuman animals being killed by us every year, I don't see how this book can change the world.

Then what's the point in writing, let alone reading, all these pages? I've struggled with this existential question for quite a while now and have settled on the following *post hoc* explanation and rationalisation: Since I've started doing philosophy, I've come in contact with people who've taught me to be a more critical thinker, to doubt authorities and myself, to question beliefs taken for granted, dissect sound sounding arguments, and to be more compassionate. It might come as a surprise that such a theoretical and seemingly dry enterprise as doing philosophy has opened my eyes to the suffering of other sentient beings. But ethical and social progress has often originated in philosophical ideas. In other words, philosophy has changed *me* and equipped me with tools that help me in attempting to change the world subsequently. And writing this book was an integral part of that enterprise.

Of course, it wasn't the amorphous enigma we call 'philosophy' that changed me, but the many people I've learned from while doing philosophy: Emmanuel Baierlé, Miloud Belkoniene, Monika Betzler, Philipp Blum, Susanne Boshammer, Davor Bodrozic, Jean Bohnert, Yves Bossart, Luzia Budmiger, Christian Budnik, Julien Bugnon, Sarah-Jane Conrad, Coralie Dorsaz, Fabian Dorsch, Patrik Engisch, Philipp Emch, Magnus Frei, Christopher Gauker, Andrea Giananti, Hans-Johann Glock, Stefanie Grüne, Andreas Heise, Silvan Imhof, Thomas

Jacobi, Rosanna Keefe, Nikola Kompa, Astrid Kottman, Elodie Malbois, Adriano Mannino, Hannes Ole Matthiessen, Michael Messerli, Res Mettler, Anne Meylan, Franziska Müller, Jacob Naïto, Jonas Rogger, Martine Nida-Rümelin, Christopher Peacocke, Klaus Petrus, Jonas Pfister, David Pitt, Tobias Rosefeldt, Mario Schärli, Charles Siewert, Michael Sollberger, Sarah Tietz, Markus Wild, David Wörner, André Wunder, Anna Zuber. To all of you philosophical friends, teachers, colleagues, and acquaintances I send out my sincerest thanks for influencing me.

Maybe my most important philosophical teacher over the last years was my supervisor Gianfranco Soldati. He saved me from prematurely quitting philosophy by giving me a job as one of his assistants at the University of Fribourg. Since then, he has opened my eyes to a different way of doing philosophy: one which focuses on true understanding for the sake of mutual progress and not on puffing yourself up for the sake of caressing your own ego; a way of doing philosophy which doesn't care about conventions, schools, and styles but about navigating the ever so puzzling seas of philosophical reflection in the most suitable vessel. I'm extremely grateful for this lesson.

During my 2016–2017 research visit at the University of Salzburg, I had the pleasure of working with Johannes Brandl. He taught me some new philosophical tricks and our joint seminar on self-consciousness was an interesting and fun experience. He also had the patience to read every word of my thesis and provided me with many good inputs for which I'm thankful. I can say with confidence that this book would be much worse without his help.

Since my time in grammar school, I had the pleasure to teach many students from diverse backgrounds and in various fields. From them I learned a lot on how to present my own and others' ideas and how to stimulate their interest, critical thinking and curiosity. I believe that the opportunity to teach other people has also helped me formulate and present the problems and ideas in this book in a clearer way. So, my thanks go to all the people I've had the opportunity to teach and influence myself.

The years writing this dissertation—which was generously funded by the Swiss National Science Foundation grant 100012_156548—were infinitely enriched by a great family of friends. The amount of support, motivation, fun, love, and inspiration I received from all these

people is an invaluable gift that defies any description. Gabrielle Christen, Kremena Diatchka, Olivier Eicher, Jürg Furrer, Maya Langenegger, Silvan Leibacher, Sebastian Leugger, Tobias Leugger, Michèle Lötscher, Nathalie Lötscher, Nicole Lötscher, Brigitte Lötscher-Müller, Josef Lötscher, Philippe Meyer, Dominik Müller, Ramon Stucki, Daniel Olivier Sutter, and my family Franziska, Valentin, Ursula, and Gisbert: I thank you, not from the bottom, but from the centre of my heart for being there.

Finally, my deepest thanks, perpetuating appreciation, and heartfelt love go to Pascale Anna, for being the wonderful and strong person she is. I dedicate this book to you!

Fribourg, April 2018

CONTENTS

PREFACE

There is not a single interesting theory
that agrees with all the known facts in its domain.

— Paul Feyerabend: *Against Method*

I promise nothing complete;
because any human thing supposed to be complete,
must for that very reason infallibly be faulty.

— Herman Melville: *Moby-Dick, or The Whale*

This book was originally supposed to be about nonhuman animals. When I started working on it as my PhD thesis in 2012, I planned on arguing that self-consciousness isn't restricted to the human species. To this end, I wanted to discuss various new insights in the fields of cognitive science, anthropology, neuroscience, and ethology. The idea was to thereby show that some nonhuman species—like bonobos, dolphins, elephants, or magpies—can become aware of, and think consciously about, themselves. But the more I read and the more I reflected, the more I asked myself the question: 'What is self-consciousness anyway?' This was the push that literally sent me down the proverbial rabbit hole.

I realised that I needed to know more about self-consciousness itself before I could seriously ask the question whether some nonhumans can become aware of themselves. So, I read up on what analytic philosophers had to say about self-consciousness. But instead of things getting clearer, I was driven even further away from my original question and closer towards the so-called 'problem of *de se* beliefs': How do we have to characterise the type of belief in which a subject thinks about herself and potentially realises that she's thinking about herself and not some other thing? It seems that there's something quite special in grasping that you're thinking about *yourself*. Just remind yourself of the mythical Narcissus seeing his own mirror image, Winnie-the-Pooh finding his own footsteps, or King Gilgamesh learning of his own mortality.

The question of *de se* thinking used to be—and still somewhat is—a hot topic in analytic philosophy of language and mind and gyrated around two influential papers by John Perry (1979) and David Lewis (1979). Trying to understand this problem and the proposed solutions, I developed the suspicion that something important was missing. The theories elaborate on the potential general logical structure of these *de se* beliefs; e.g. how they differ from and relate to beliefs about other things. But they didn't explain how subjects come to understand that they're thinking about *themselves* in the first place. And this curious way of getting in touch with yourself is exactly what makes self-consciousness special. If I wanted to understand how nonhumans can become aware of themselves, I had to find answers to the question: What's going on when a subject thinks about herself in this peculiar *de se* way? What's special about thinking about *yourself*?

My focus thus shifted from the enigmatic question concerning the nature of self-consciousness to the more concise, narrow, and previously staked out one of these unique *de se* thoughts: mental states which are always and necessarily about whoever is thinking them. And by doing so, I soon came to realise that there's a certain structural reflexivity which I wanted to understand better. When I think that I'm typing, this thought seems to be about myself no matter what. There's no way I'm thinking about some other person when I'm thinking in the *de se* way. How's that possible?

Unfortunately, the usual suspects in the literature didn't seem to address what perplexed me the most—that subjects can grasp that they're thinking about themselves *at all*. They rather started from the premise that this just occurs and now has to be distinguished from other types of thinking. Hence, I became somewhat frustrated with the overall technicality and obsession with theoretical details that's prevalent in the academic discussion of the problem at hand. I asked myself: Can't we explain the distinctiveness of *de se* thinking from a less technical point of view? Can't we show the general problems of the pertinent academic literature without struggling with all the small quirks and minute differences between the various theories on offer?

Now, as you might have guessed, this book is an attempt to give and justify an affirmative answer to both these questions. As such, the overall structure of the book is more akin to a philosophical journey than the more usual series of arguments, reconstructions, objections, and

minute refinements. In a way it represents my own odyssey through the topic of *de se* thinking and my various encounters with authors, ideas, and crucial insights. The aim is thereby to provide the reader with an impression of how the problem initially unfolded, to show which theoretical foci and decisions lead to which approaches, and to illuminate the problems that these approaches face and have to overcome. And, of course, to sketch out and defend the solution I think is most worthwhile in the end.

Despite the complexity of the debate and the many articles that have been written about technical issues, minute improvements, and new small differentiations, I set myself the goal of illuminating this very narrow and seemingly insignificant problem in a way which tries to remain as non-technical as possible. Therefore, I decided to sidestep certain questions of detail which distinguish the various theories on offer. Rather, the goal was to identify general *strategies* which are employed in the debate and show the general problems such strategies face. Furthermore, I wanted to give an impression of how we even end up with these strategies, where they originate from and why they were deemed worthwhile or doomed. This more pedagogical approach might be somewhat unusual. However, it follows from my conviction that important philosophical points can be made perfectly well on such a general level and that great attention to technical details brings rigorousness on the cost of losing sight of the original question.

By looking at the bigger picture, I had to abstract from the finesse of the accounts that have been developed over the last 40 years. But, of course, we'll naturally encounter several specific theories and quotes from important authors despite my subscription to a more relaxed and less academic style. These encounters will give us the opportunity to get a better picture of what we're dealing with and also learn some important lessons. Moreover, they're aimed to indicate that I didn't just pull the discussed strategies out of my hat. Admittedly, the conversation with these other philosophical theories will be somewhat naïve. I intentionally gloss over some of the details of these accounts. And this again sometimes has the air of an exercise of building strawpeople. However, I invite you to rather understand my discussion of the pertinent literature as a way of illustrating and motivating the general strategies that emerge in the philosophical exchange. My interest is decisively not in showing the infeasibility of one specific theory or

another. I'm much more interested in kinds of advantages and problems of these more general structures. As a result, I don't intend to dismantle opposing theories for good but I rather try to clearly point out their weak spots and problematic aspects—which can then potentially be overcome by their proponents.

The general dialectic and methodology of the book thus amounts to the following: We start from the simple fact that subjects oftentimes think about themselves in the special *de se* way. This way of thinking has some quite general characteristics which distinguish it from other types of thinking. We can identify and illuminate these characteristics by helping ourselves to the multitude of philosophical theories and insights which have been produced over the last two millennia—where we'll gladly focus on the last fifty or sixty years.

In the course of this, we'll also encounter strategies which have been executed by different philosophers in order to solve the problem of *de se* thinking and to explain what's special about our ability to think about ourselves. These strategies again need to be outlined in some way or other. It's sensible to do this by taking some representative theory or author as a template against which we can develop the strategy further and indicate its most important insights and problems. Such a template and gross generalisation comes with the problem of not doing full justice to the advocate of a certain theory. Nonetheless, I'm confident that there's some merit to paying this price.

Most importantly, the lessons we learn from this exposition of attempts to solve the problem at hand can ultimately be used to illustrate, compare, and contrast the solution I prefer. Our journey through the universe of unsuccessful attempts will teach us not only what doesn't work, but also what's essential to any good account of *de se* thinking. In this way, we'll end up with the account that I will try to defend and make as compelling as possible. My own solution to the problem takes up an old insight that has been quite prominent in phenomenology and so-called 'existential' philosophy. Building upon this, the book tries to integrate it with the learned lessons which originate in the more analytic style of doing philosophy from which we started.

What's this solution? It starts from describing subjects as the origins of their world: they perceive and act from a certain privileged position in space which they themselves occupy. After all, much of what we see and do is put into relation with ourselves. We see the snow-covered

mountains that we want to climb in summer. We see our friend that we want to hug. Now, getting into touch with the origin of our world is tantamount to getting into touch with ourselves. And thus, *de se* thinking finds its beginning in the fact that we're first and foremost origins. We are living bodies, perceiving and acting subjects. This also somewhat explains the cryptic title of the book. The fact that subjects are origins of thought and action is the key to understanding the peculiarity of *de se* mental states.

As already mentioned, this idea is in no way new. It can probably be traced far back to idealism and transcendental philosophy influenced by Kant or Hegel and of course to phenomenologists and existential philosophers such as Husserl, Sartre, and Merleau-Ponty. Recently, similar accounts have been developed in the philosophy of mind under the headings of pragmatism, embodied cognition, and enactivism. These are all sources that inspire and illuminate my own account. This is why I don't intend to claim to have invented a brand new and previously unknown idea. Rather, my goal is to put certain perspectives, strategies and problems into relation to each other which have previously been mostly discussed independently. As such, the approach defended in the book is much more a previously overlooked solution to the problem of *de se* thinking than a new theory altogether. On the one hand, it shows how we have to arrive at this more phenomenological insight even if we start from the dry desert of analytic philosophy. On the other hand, it incorporates this insight into the analytic debate in a way which hopefully enriches and irrigates it.

Now, after the scope and style of the book have been elucidated, is the opportunity to give a more detailed overview over all the chapters, the general arguments, and important claims that will define its content. The main story is divided into four chapters. In the first chapter *Being in the mirror*, I will introduce the phenomenon of *de se* thinking and develop a conceptual and methodological framework which will be used in the remainder of the book. The history of philosophy is intimately tied with the ability to think about ourselves. From Plato's quest for self-knowledge to Kant's necessity of the 'I think'; from Descartes's *Cogito* argument to Camus's Sisyphus. Several of these sources provide us with a better understanding of the ability to think about ourselves. In the course of this voyage, we'll come across a number of characteristics that are connected to *de se* thinking: its semantic reflexivity, the con-

nection to self-knowledge, or its necessity for intentional action and behaviour. One of the most important methodological desiderata follows from this list: Any proper account of first-person thinking should be capable of accommodating these features in some way or other.

From this general introduction and initial theoretical setup we then start to test the waters and look at some of the more prominent approaches to *de se* thinking. How do they emerge? What's their main general claim? What are their advantages and problems? And can they do justice to the characteristics of *de se* thinking? We start this survey in the chapter *Divide and conquer* where we'll familiarise ourselves with one influential family of approaches. This family is heavily influenced by the idea that our mental states can be characterised by some proposition—a kind of abstract picture of how the world could be—which we're related to in believing or desiring. For instance, a subject believing that it's going to rain on Sunday entertains a 'picture' of the world where it rains on that particular day. Accordingly, her belief is true if that picture is accurate and it indeed rains on the day in question. Unfortunately, such a simple picture isn't suitable for our thoughts about ourselves. Due to the innate reflexivity and dependency on the context—one and the same *de se* belief corresponds to quite different propositions when entertained by different subjects—we'll shift the focus to so-called *two-dimensional* theories. These are designed in the spirit of the propositional family but with some extra features. Very generally, they hold that our beliefs have to be analysed in terms of two distinct logical steps: one of them being sensitive to the context of the mental state. The chapter will explore three different ways of developing this idea: the linguistic, the conceptual, and the functional approach. I will illuminate the benefits and problems of these different strategies and ultimately argue that the most important puzzle piece is missing. They all fail to explain how subjects can come to realise that they're thinking about *themselves*.

What do you do when your hopes for salvation are thwarted by one camp? You turn to the other. In the subsequent chapter, *Back to the primitive*, we'll discuss the main opponent to the idea that mental states should be characterised by some proposition. Rather, this strategy claims that we should focus on the kind of *property* that's in play in a given belief, supposition, desire, or intention. So, when I think that I'm sleepy, we should say that I simply ascribe the prop-

erty of being sleepy to myself. And if I think that you're sleepy I'm simply ascribing that same property to you. This property theory of *de se* thinking was designed specifically with our thoughts about ourselves in mind. It even goes as far as claiming that any mental state we have is ultimately a thought about ourselves. Why? Because in any case, we're ascribing a very complex property to ourselves. Even if you believe something about the weather, you're ultimately believing that *you*'re living in a world where it's windy, rainy, warm, or cold. We'll see that this way of painting the conceptual landscape comes with certain problems. Most importantly, we run the risk of losing sight of what makes proper *de se* thinking special. While the propositional theories weren't able to fit the *de se* way of thinking into their picture, this new property theory overshoots the target. I'll argue that proper *de se* thinking is related to a very special kind of self-ascription. More precisely, we have to identify a type of self-ascription which is primitive—which doesn't depend on some other kind of knowledge or belief. This way of self-ascribing a property is what forms the basis of *de se* thinking and has to be sharply contrasted from other types of thinking. Unfortunately, the classical property theory doesn't deliver this result. Hence, we'll have to abandon it for a more feasible alternative.

The lessons from the two previous chapters set up the story for the grand finale. After having argued that two-dimensional theories can't explain how subjects come to think of themselves in the peculiar *de se* way, our examination of the property theory taught us the necessity of some primitive form of self-ascription. Naturally, if we accept the necessity of this primitive kind of self-ascription, we better explain its nature and how it can serve as our required basis of *de se* thinking. This is the goal of the fourth and final chapter *Origins*. Here, I'll explicate the proposed solution and defend it. The argument starts with a characterisation of what's called *egocentric space*. Subjects think about the objects in their world primarily from their own first personal perspective. And this amounts to them putting these things in relation to themselves. For instance, I'm thinking of the glass as being to the left of *me* or I realise that *I* can't see Mount Fuji from here. This kind of thinking is intimately tied to a subject's capabilities for action. Basically, the argument is that for a subject to be capable of interacting with all the glasses, mountains, people, or animals around her, she has to think about them egocentrically.

How does this relate to primitive self-ascription or *de se* thinking? This is where the concept of the origin comes into play. In the chapter I'll argue that a subject takes up a very special position in her egocentric thinking. She's the phenomenological and behavioural centre of her world. She occupies the central place from where she experiences the world around her and herself. And at the same time she acts upon the objects in the world from that centre. Following the title of the chapter, I call this the *origin*—which is further characterised by the concept of the lived body. We're all subjects who engage with the world through our lived body—the thing with which we see, feel, walk, talk, cry, love, or stumble. In the argument this lived body plays a crucial role since it enables these primitive self-ascriptions which we're after. By experiencing the world through the lived body we at once take certain properties to be instantiated in ourselves. We wouldn't try to get up if we didn't think that we're lying in bed. We wouldn't grab a beer if we didn't think we're thirsty—well, we could, but you get the point. From these arguments and considerations we can develop the theory that *de se* thinking is always rooted in some ascription of a property to the lived body. The lived body is inherently—through it being the origin of our thought and action—given to a subject as her own. By ascribing a property to that lived body it inherits this first-personal aspect and explains why *de se* thinking is so peculiar.

Once the concepts of egocentric space, the origin, the lived body, and the favoured theory of *de se* thinking is adequately explained and its necessity properly argued for, we'll connect it with the starting point of our journey. I'll explain how such a way of framing things does justice to the characteristics of our ability to think about ourselves. In that context we'll see that it's not so easy as we might expect and that—following the mottos for this book—more work needs to be done. We'll see how the defended account deals with various types of *de se* thinking and, more importantly, some problematic objections. This will show us both the virtues of the account and its pitfalls and limits.

These four chapters form the principal line of argument. We can summarise it in the following way: *De se* thinking has several characteristic features which aren't present in all instances of thinking about yourself but are at least potentially realised. As such, any feasible account needs to explain the potential for these features. Neither the two-dimensional accounts—stemming from the idea that mental states

can be characterised using the notion of a proposition—nor the property theory—claiming that we self-ascribe a property in thinking—do full justice to the phenomenon at hand. Instead, we have to take the concept of primitive self-ascription as the basis for all *de se* thinking. Primitive self-ascription, in turn, is ascription of a property to the lived body. The lived body is constituted through a subject's assumed possibilities of interaction with the world. Therefore, only subjects as lived bodies are capable of thinking about themselves in the *de se* way. This is what's required to grasp that you're thinking about *yourself*.

As you will see, the arguments and dialectic of the book aren't always as straightforward as one would expect on the basis of this concise précis of the overall argument. Rather, it oftentimes takes on an almost meandering way of exposition. This is partly deliberate because it represents the way the academic discourse has developed. Original thought doesn't always follow the straight line of hindsight and to understand the development of certain answers it might be necessary to tailgate the twisting and turning path. Otherwise, we'll often be left wondering: 'How did we get here?' Understanding certain theoretical choices and failures is easier when we grasp their emergence. Because of this swerving dialectic it might be helpful to have a list of important claims that are endorsed throughout the book:

- Thinking about oneself in the *de se* way doesn't require a self-representation.
- There's both unconscious and nonconceptual ways of *de se* thinking and thought in general.
- The ability to think about oneself consciously and explicitly has its origin in cognitively simple abilities such as the unconscious grasp of the lived body.
- Thinking in the *de se* way isn't the same as self-consciousness.
- Thinking is possible without language.
- Subjects are constituted through their lived bodies.
- In the foundational cases, the lived body is not represented in thought.
- Thinking in the *de se* way is possible due to the subject's ability to primitively self-ascribe properties.

Moreover, there are some claims which you might think I endorse but on which I choose to remain neutral or even explicitly don't defend. The following list collects some of these claims:

- Self-knowledge is special.
- Self-knowledge isn't special.
- Reasons for action are always facts.
- Self-representation is impossible.
- There's no adequate two-dimensional account of *de se* thinking.

I hope that these lists of claims and non-claims will help to illuminate and put into perspective the arguments in this book. However, there is another part of the book which I hope is illuminating to the more academically inclined reader: The main part of the book is followed by an appendix. Sometimes, during the main story, you'll come across references like this: (A.3.2). As mentioned, part of the goal of this book is to make it as accessible as possible—even to people outside of the debate or academic analytic philosophy in general. This is why it uses almost no footnotes and introduces even fairly well-known philosophical concepts in an undemanding manner. Sometimes, however, I feel the urge or the necessity to say something more about a topic without interrupting the flow of the argument. The appendix is the arena for these conceptual and argumentative excursions. This is where I might clarify certain ideas in a more formal or technical way in order to make my claims more precise. I also point towards interesting or important relations with the academic literature. As the main story uses only few references to other people's work, it's sometimes necessary to show some understanding of and sensitivity to the work done by others. So, whenever you see such a reference and you're interested in what I have to say about it check the referenced section of the appendix for some additional information.

Finally, I'll use the following typographical conventions throughout the book: Whenever I'm talking about a sentence that expresses a thought, intentional state, or mental state, I will put the sentence in quotation marks. On the other hand, when I want to talk about the underlying expressed thought without committing myself to any kind of linguistic connotation I'll put the (closest) sentence that is usually used to report the thought in italics. And whenever I'm talking about a proposition I'll put the proposition into chevrons. Sometimes

I'll write out the propositions more colloquially and sometimes more formally. So, I'll talk as if the sentence 'The leaf is green' reports the subject's thought *The leaf is green*. The subject believes that the leaf is green. And the proposition that's expressed by the sentence is <that the leaf is green> or <the leaf, being green>. Of course, I'll use italics for emphasis throughout the book as well.

1

BEING IN THE MIRROR

> Here Narcissus, tired of hunting and the heated noon, lay down, attracted by the peaceful solitudes and by the glassy spring. There as he stooped to quench his thirst another thirst increased. While he is drinking he beholds himself reflected in the mirrored pool—and loves; loves an imagined body which contains no substance, for he deems the mirrored shade a thing of life to love. He cannot move, for so he marvels at himself, and lies with countenance unchanged, as if indeed a statue carved of Parian marble.
>
> Ovid: *Metamorphoses*: Book 3, 407

Self-consciousness is oftentimes used as a defining criterion and a prime example of what makes us human—almost our *conditio humana*. In this manner, we elevate ourselves from other beings through our supposedly unique ability to consciously think about ourselves. However, right there in our midst, Ovid's Narcissus is an epitome of the tragedy that often accompanies self-consciousness. The myth demonstrates the deep chasm that comes with the possibility to consciously think about oneself. Indeed, only once Narcissus is self-consciously marvelling at himself does he realise his terrible predicament. If only Narcissus wouldn't have known himself, he might not have suffered the direful fate of needing to commit suicide to escape from his prison of self-consciousness.

This is but one example where our ability to self-consciously think about ourselves comes at a price. Others are aplenty: we think ourselves too stupid or smarter than the rest and even sometimes wish we were someone else. Without being able to consciously think about ourselves

in this way, we wouldn't suffer from being self-conscious about our weight, our desires, our actions, our personalities. Hence, even if our ability to consciously think about ourselves were a gift for our human existence, it's also a bane—as so often: there ain't no such thing as a free lunch.

Sure enough, self-consciousness can also be a virtue. Only through our ability to consciously think about ourselves are we capable of building complex societies—founded on justice, laws, and politics—and grasping our own role and position in such a society. Only self-consciousness allows reflection on what kind of person we want to be, what goals we do and don't want to pursue in life. And finally, it's tightly connected to our understanding of what it is to be someone else. The possibility of serious and earnest empathy might only arise in concert with our appreciation of ourselves as vulnerable beings in the midst of other vulnerable beings—an impossible achievement without self-conscious thought about oneself. So, it's no miracle why we hold our ability to think about ourselves in such high regard despite some of its undesirable consequences.

But self-consciousness doesn't just imply some practical adversities. It also produces the most serious philosophical problems. For instance, without self-consciousness, we would be incapable of grasping our cosmic insignificance and there wouldn't be the dreaded possibility of suicide—a perspective on self-consciousness that's perhaps most pronounced in what's commonly called *existentialist* philosophy. Of course, this doesn't imply that every self-conscious subject is in danger of killing itself. Even Sisyphus, fully aware of his futile purpose in life, doesn't contemplate suicide—*au contraire*, we must imagine him to be a happy person (Camus 1955). This is because self-consciousness doesn't just present us with the vacuity of our own earthly lives; it also allows us to give our lives purpose and meaning. Nonetheless, without self-consciousness the fundamental question of the meaning of life would never have been contemplated.

But what is self-consciousness? What's special about it? How do we distinguish thinking about oneself from other kinds of thinking? These questions form the cornerstones of this book. It's an attempt to understand what's special about our ability to think about ourselves and explain what distinguishes it from thinking about other things in the world. I want to know what's involved in our ability to think about

ourselves and how we can best explain the peculiarities of these special kinds of thought.

Let's approach these delicate questions by first distinguishing different varieties of thinking about oneself. That's important because we sometimes aren't remotely aware of the fact that we're thinking about ourselves. Think of the first woman in space: Valentina Tereshkova. Imagine she's reading the newspaper headline: 'First woman in space just turned 80'. As a consequence, she thinks to herself: *the first woman in space just turned 80*. Now imagine further that she isn't at the height of her cognitive abilities early in the morning and doesn't remember that she *herself* was the first woman in space. So, in one sense, she's thinking about herself. After all, the person she's thinking about is Valentina Tereshkova—herself. But in another sense she isn't really thinking about herself because she doesn't *realise* that she herself is the first woman in space—she's not aware of the fact that she's thinking about herself.

Or take another well-known case: you got lost in the woods and are running in circles trying to find a way out. You come across your own footprints in the mud. And now you mistakenly think that the footprints were made by someone else and hence think to yourself: *Someone was here*. Of course, it's yourself you're thereby thinking of. After all, the mysterious 'someone' is you. But again, you don't realise this. Rather, the fact that you continue to follow the footprints shows that you're oblivious to the fact that you're running in circles.

Winnie-the-Pooh provides an illustrating example of this obliviousness. He and Piglet find themselves in the predicament of mistaking their own footprints for someone else's when they attempt to find a woozle by following its footprints (Milne 1926). What they don't realise is that they're following their own footprints and mistakingly believe they belong to a woozle. So, when they think about the inexistent woozle—e.g. when Winnie-the-Pooh tells Piglet 'The woozle went around that tree'—they're really thinking about themselves without realising it. It's Winnie-the-Pooh who went around that tree.

On the basis of these examples we can introduce an important distinction between three different varieties of thinking about oneself. The first kind is one which doesn't imply that the object of your thought is in fact yourself—it's a mere stroke of luck. When you think *Someone was here*, it theoretically could've been anyone. Maybe in a different

possible world the footprints were in fact produced by a woozle. This kind of thinking can be characterised as thinking about *something or other* having certain properties. And as luck would have it, it happens to be yourself who instantiates the relevant properties. In the same sense Valentina Tereshkova's instance of thinking *The first woman in space just turned 80* is of that type because someone other than her could've been the first woman in space. Maybe there's a possible world where you're the first woman in space and so in that world she would've been thinking about you. This kind of thinking is usually called *de dicto* thinking about oneself. You're thinking about something or other which happens to be you without necessarily realising that it's you you're thinking of.

What's important for this kind of thinking is that the way you're thinking about yourself is via some description or other. When Valentina Tereshkova thinks of herself as 'the first woman in space' she's using a description that could potentially be satisfied by any object and which only happens to be satisfied uniquely by her in our possible world. And—other things being equal—whether she's thinking about herself in thinking *The first woman in space just turned 80* or someone else doesn't change the nature of her thinking. In any of these cases she's identifying the thing she's thinking about by the description 'the first woman in space' which can pick out a number of things.

However, sometimes you think about yourself in a more direct and specific way without at the same time realising that it's you you're thinking of. The case of the amnesiac illustrates this in a formidable way. Imagine Alpha, an amnesiac, standing in front of a mirror, seeing her reflection and thinking at first to herself *That woman is tall*. She doesn't realise that she's standing in front of a mirror and thus doesn't realise that she's thinking about herself. For the moment, we have a classical case of *de dicto* thinking about oneself. Now Alpha isn't your usual amnesiac. She's very capable of recognising faces and people and putting the right names to the right faces. So, after a moment, she realises that the woman in the mirror is Alpha—an old acquaintance she somehow remembers—but she doesn't recognise herself. Subsequently, her thought changes and she now thinks *Alpha is tall*.

This new kind of thinking is different from *de dicto* thinking because she now doesn't think about something or other which satisfies some description. She's now thinking about something very specific: a par-

ticular thing which couldn't be something else. Alpha is Alpha in all possible worlds, whereas the woman in the mirror could be someone else in a different scenario. This second kind of thinking is usually called *de re* thinking about oneself. In these cases we're thinking of a specific particular thing—a *res*—and not just something that satisfies a certain description—a *dictum*. However, as the case of our amnesiac Alpha shows, thinking *de re* about oneself still doesn't imply that the thinking subject realises that it's her she's thinking of.

So, we finally come to the third variety, which got Narcissus into trouble and many philosophers alike. This variety of thinking about oneself carries an intimate connection to the entertaining subject and above all entails some minute grasp of who's the thought's *dramatis personæ*: yourself. Valentina Tereshkova doesn't suffer from amnesia and she knows that she herself just turned 80. In other words, she thinks something along the lines of *I just turned 80*. And in virtue of thinking about herself in that way she's aware of the object of her thought. She grasps that it's herself she's thinking of.

In the same way you're aware of who's doing the reading while you're reading this book: it's yourself. And when you're cooking you can be aware of who's doing the cooking: it's yourself. This third variety of thinking about oneself is usually called *de se* thinking and interests us the most. You're no longer merely thinking about an object that fits a certain description—as in the case of mere *de dicto* thinking. And you're no longer just thinking about a specific particular thing without necessarily realising that the thing is you—as in the case of mere *de re* thinking. You're now thinking explicitly about yourself: Congratulations!

It's important to note, however, that *de se* thinking is only typically a conscious activity. Self-consciousness is just the most apparent and—for us—easily graspable form of *de se* thinking. However, there are many cases where you're thinking about yourself in the *de se* way without being aware of that fact. For instance, seeing the tennis ball approach you, you're thinking about the tennis ball in relation to you. Let's say that you're trying to hit a forehand winner. This requires you to position yourself in a specific way relative to the ball. So, while you're consciously focused on the ball quickly flying your way you're at the same time thinking about the movement of your feet, your arm, your eyeballs, your knees. All this is necessary for a successful winner

but none of it needs to be conscious. In these cases you're thinking *de se* without these thoughts becoming conscious. Similarly, saying that *de se* thinking comes with a certain grasp or awareness of who you're thinking of merely means to convey that you can't think about something else in that particular way.

We've now marked out the topic of this book on the basis of a first conceptual distinction between three different varieties of thinking about oneself. What we're interested in is *de se* thinking—but that's not enough. We want to know what's special about this kind of thinking and we've already witnessed in passing some of the features of our ability to think about ourselves. I'll now discuss these peculiarities of *de se* thinking in a bit more detail (A.1.1).

1.1 THINKING ABOUT YOURSELF

One of the more basic features of *de se* thoughts is that they're always *about* the subject that's thinking the thought. Take an example: In one sense, Valentina Tereshkova thinks about herself in the same way when she believes *I just turned 80* as she's thinking about Steffi Graf in believing *Steffi Graf won 22 Grand Slam singles titles*. What do I mean by 'the same way' here? Well, the question 'Who are you thinking about?' would be answered with 'Myself, Valentina Tereshkova' in the former case and with 'Steffi Graf, the tennis player' in the latter. What a thought is *about* is oftentimes called the 'intentional object' of the thought. And in the case of *de se* states, the intentional object is always the thinking subject (A.1.2).

With respect to the intentional object, *de se* thinking is different from *de dicto* thinking because the latter isn't always about the thinking subject. Of course, when Valentina Tereshkova believes *The first woman in space just turned 80*, she's also thinking about herself. And hence, she's the intentional object of that instance of thinking. However, not all instances of *de dicto* thinking are like this. Someone else might have been the first woman in space and then the intentional object of Valentina Tereshkova's *de dicto* belief would've been that other woman. So, a subject can think *de dicto* about a variety of things and hence, her thinking can have a wide variety of intentional objects.

Something similar applies to cases of *de re* thinking. We can think *de re* about a whole lot of things. Sure, whenever a subject thinks *de*

re about herself—as in Alpha's belief *Alpha is tall*—she can't fail to be the intentional object of her thinking. Additionally, that's true in all possible worlds—in contrast to thinking *de dicto* about oneself. Alpha always thinks about Alpha, i.e. herself, in virtue of believing *Alpha is tall*. But Valentina Tereshkova doesn't always think about herself in virtue of believing *The first woman in space just turned 8*0 because in some other possible world you could've been the first woman in space.

Nonetheless, *de re* thinking doesn't imply that the thinking subject is always the intentional object because we can think about many things in a *de re* way: When Alpha believes *Venus is bright tonight* she's thinking *de re* about something; but it ain't herself. She's thinking about the planet in the sky and so in this case Venus is the intentional object of her belief.

We've seen that neither *de dicto* nor *de re* thoughts have the feature of always being about the thinking subject. That's unsurprising. As we saw, we can think *de re* about all kinds of things: apples, Steffi Graf, or me. And the same applies to thinking *de dicto*. But every subject can only think about one thing in the *de se* way: herself. This makes thinking in the *de se* way special: it's always about ourselves. Hence, the intentional object of *de se* thinking is always, i.e. necessarily, the thinking subject. When you're thinking about the position of your feet before hitting a winner, you can't fail to think about yourself. And equally, when Narcissus contemplates suicide he can only think about himself and nobody else.

A closely related second feature of *de se* thoughts concerns the question: What needs to be the case for my *de se* thinking to be accurate? We often want to know what needs to be the case for a belief to be *true*, a desire to be *fulfilled*, an order to be *executed*. In other words, we're often interested in the conditions of satisfaction of our mental states and attitudes. And in the case of *de se* thinking the conditions of satisfaction always depend on the thinking subject in a specific way. This is, of course, a rather direct consequence of the fact that the thinking subject's always the intentional object of *de se* thinking.

To illustrate this point let's look at the belief *I am tall*. We can ask: Under which conditions is that belief *true*? Well, there's no direct answer because that depends on who entertains the belief. Is it Alpha, or Beta, or Steffi Graf? If Alpha believes that she herself is tall, then the belief is true if Alpha is tall. And if Beta believes that she herself is tall,

then the belief is true if Beta is tall. Hence, what makes a *de se* belief true depends on who's entertaining it. But what does it mean that the belief's truth-conditions 'depend' on who's entertaining it?

In general we might say that the conditions of satisfaction of a mental state are given by whatever *proposition* tells us what needs to be the case for the belief to be true, the desire to be fulfilled, the order to be executed. But what's a proposition and how does it help us? Let me give an example of the role of a proposition and explain after: When a subject believes that Sydney is the capital of Australia she has in mind a certain possible way the world could be. This way is characterised by having Sydney as the capital of Australia. Now, a proposition can be understood just as a possible way the world could be. And since it's possible that Sydney were the capital of Australia there's a proposition which 'tells' us that. So, we can use this proposition—an abstract entity which tells us how things could be—in order to characterise the subject's belief.

There's a whole bunch of theories on what propositions are exactly, e.g. what their metaphysical, semantic, logical, and syntactic features are. This dispute needn't bother us for the moment. For the illustrative purpose of this excursion it'll be enough to take what I'll call the Propophile's theory of propositions as an example. According to the Propophile a proposition is a structured abstract entity that tells us how things could be. Furthermore, it's a bearer of a truth-value in a certain absolute sense. It can be either true or false—at least in the most simple and accessible case of standard bivalent logic. Propositions are true absolutely in the sense that it's once and for all determined in which possible worlds they're true and in which they're false—there's no middle ground.

A possible world is a world pretty much like the one we're living in with small or large differences. For instance, there's a possible world which is identical to our actual one—it includes all our laws of nature, the planet Venus, people who believe in God, this book you're currently reading, the Sombrero Galaxy M104, and so on—save for the fact that I'm left-handed. And hence, there are myriads of possible worlds. Every tiny change in how things could possibly be gives rise to a new possible world. In some of these worlds, the laws of nature are altered, in others there might be no laws at all. Some philosophers think that possible worlds are merely abstract products of our ima-

gination, useful to think about what's possible and what's impossible; others think that they're just as real as the world we're actually living in. Again, we don't let ourselves get held up by these discussions. We can use the talk about possible worlds just to talk about alternative scenarios how things could be or turn out. It doesn't matter for us whether these scenarios merely play themselves out in our heads or actually 'out there'. And that's just fine.

Back to propositions: More important for our topic of *de se* thinking is the following feature that Propophiles have in mind: If a proposition is true in a specific possible world, it's true for everyone in that world. Pluralism and relativism about truth aside, propositions, so we're told, can't be true for you and false for me (A.1.3). How does that work exactly? Take the belief that Sydney is the capital of Australia as an example. Here's the proposition that interests us:

(1) <Sydney, being the capital of Australia>

As a matter of fact that proposition is false in our world. Sydney isn't the capital of Australia, Canberra is. And that fact doesn't change whether you or I believe that Sydney is the capital of Australia. How can a state in the world make an abstract entity like a proposition true or false? We can be happy with the metaphorical way of speaking in which a proposition is like a picture—something made popular for a while by the young Ludwig Wittgenstein (1922). Pictures might be accurate or inaccurate representations of reality. And propositions belong to the realm of hyperrealism: the more accurate, the better. The picture that's painted by the proposition (1) doesn't represent our reality accurately at all. Hence, the proposition is false in our world. In a different world (1) might've been true, but not in ours.

Classically, Propophiles have used propositions in order to characterise and individuate mental states. If you want to know what a subject believes, you just need to know what proposition she thereby holds true. We can ask the metaphorical question: 'What picture of reality is she holding in her mind?' Take as an example the case of believing that Sydney is the capital of Australia. We can now characterise your belief using the proposition (1) and *ipso facto* distinguish it from other beliefs you could have. The belief that Sydney is the capital of Australia is different from the belief that Canberra is the capital of Australia because the former is characterised by the proposition (1) while the latter is

characterised by the distinct proposition <Canberra, being the capital of Australia> (A.1.4).

Interestingly, propositions don't just help us tidy up our mental lives in telling us what subjects believe. They also tell us what needs to be the case for the relevant belief to be true. For instance, if Alpha believes that Sydney is the capital of Australia, we know that she therefore holds the proposition (1) to be true. That proposition matches the picture of reality she has in mind. So, if Sydney has the property of being the capital of Australia in our world, the proposition is true and Alpha's belief likewise. However, the city with the characteristic opera house doesn't have the property in question. Hence, (1) is false and any belief that's characterised using that proposition is likewise false.

So, propositions can be used to give us the conditions of satisfaction of our beliefs, desires, or hopes. Of course, we can also use propositions to characterise our *de se* thoughts. For instance, the proposition <Alpha, being tall> tells us what needs to be the case for Alpha's belief *I am tall* to be true. She believes truly just in case Alpha is tall. And that's to say that she believes truly just in case Alpha has the property of being tall: exactly the picture that's painted by the relevant proposition.

What's special about *de se* thinking is that we need to know who's thinking in order to know which proposition is relevant to determine the conditions of satisfaction. The belief *I am tall* can be entertained by many different subjects. Thus, we need a different proposition that tells us the truth-conditions of the *de se* belief for each case. If Alpha believes *I am tall*, the relevant proposition is <Alpha, being tall>. Her belief is true just in case Alpha, the believer, has the property of being tall. And likewise, if Beta believes *I am tall*, the relevant proposition is a different one: <Beta, being tall>. So, her belief is true just in case Beta, now the believer, has the property of being tall. We can generalise from these observations and describe our second characteristic feature of *de se* thinking—that the conditions of satisfaction of *de se* thoughts depend on who's thinking—in the following way: A subject's *de se* thought of the form *I am F*, where *F* is any property you want, is satisfied just in case the thinking subject instantiates the property in question.

There's an interesting contrast here to cases of *de re* and *de dicto* thinking. We saw that it doesn't matter who's believing that Sydney is the capital of Australia in the Propophile's picture: the proposition (1) al-

ways determines what needs to be the case for the belief to be true or false. In contrast to the *de se* case we don't find any variability in the case of belief *de re* and *de dicto* between the specific entertained belief and the proposition that gives the conditions of satisfaction. But *de se* thinking comes with such variability due to its crucial reflexivity. Accordingly, *de se* thinking needs a special treatment in order to get from the thought to the relevant proposition. As we'll see this has given a lot of people headaches.

These are then the two first features of *de se* thinking which we need to keep in the back of our minds when we answer the question: What are *de se* thoughts and how can we account for them? On the one hand, we have the seemingly trivial feature that we're always thinking about ourselves when we're thinking in the *de se* way. This might sound like an insignificant feature but it's one of the driving insights in developing a theory of *de se* thinking. Necessity—such as the fact that we're necessarily the intentional object of our *de se* thoughts—belongs to philosophers' most priced toys and is certainly worth considering. On the other hand, we saw that the conditions of satisfaction of our thoughts about ourselves depend in a systematic way on the thinking subject. Whenever a subject entertains some *de se* thought the thinking subject will be part of what determines the conditions of satisfaction of that episode of thinking. That's something which distinguishes our *de se* thoughts radically from thoughts about things in the 'outside world', which don't exhibit this kind of dependance. We can see that the two features are closely related: It's because of the fact that the thinking subject is the intentional object of *de se* attitudes that the conditions of satisfaction depend on who's thinking the thought in question.

1.2 A VACCINATION FOR THOUGHTS

The next feature I'll illuminate concerns the problem that the intentional object of our thought sometimes doesn't conform to the actual object which brought the mental state about. We assume that we're thinking about something but we actually misidentified that thing and are instead thinking about something else. Interestingly, *de se* thinking sometimes enjoys a special kind of immunity to this identificational error in thinking. What's that supposed to mean? When a subject thinks about a specific thing—for instance the tree in her garden—she can

go wrong with regard to what brought that mental state about. There might be no tree, or what she takes to be a tree is in fact a vividly hallucinated friend. In such a case we can say that the intentional object of her thought is the tree in her garden that she's having a visual experience of. After all, that's what her thought 'intends' to be about.

However, she can't base her belief on the tree's existence because there's no such tree. What actually made the subject think that there's a tree in the garden is the presence of her friend—the source of her perceptual experience. It's just that through the use of hallucinogenic drugs her friend appears to the subject in the disguise of a tree. Despite the fact that the subject's belief is about the tree in the garden with its beautiful red apples, there isn't such a tree out there. She misidentified the object of her belief as a tree instead of a friend. What she's intending to think about just isn't there. In such a case we could say that no thing in the world conforms to the intentional object of her thinking. Her thought isn't actually about the thing it was meant to be about.

This kind of error creeps up on many of our intentional states. We're often wrong about what we thought we identified in thinking. The subject in our example seemed to see a tree in the garden but she *misidentified* what her thought was about. The thought was brought about by her friend and not by the tree in question. Another standard case is the following: Beta might see herself in the mirror and think that she's tall while in fact the reflection is of her twin sister Gamma via a complex system of mirrors. While she's the intentional object of her thinking Beta's belief is based on the presence of her twin sister Gamma because she's the origin of her perceptual experience; after all, it's a reflection of Gamma. These examples show that we're generally susceptible to misidentifying the object of our thoughts. Sometimes our aim isn't as good as we wish—a fact that powers many skeptical scenarios and challenges.

Wittgenstein is among the first in the 20th century to have made some more or less systematic study of this identificational error in thinking. In his *The Blue and Brown Books* (1958), he describes that our *de se* thoughts are peculiar with respect to the possibility of misidentifying the intentional object. He contends that such an error isn't possible in some cases of thinking about ourselves:

> There are two different cases in the use of the word 'I' (or 'my') which I might call 'the use as object' and 'the use as subject'. Examples of the first kind of use are these: 'My arm is broken', 'I have grown six inches', 'I have a bump on my forehead', 'The wind blows my hair about'. Examples of the second kind are: '*I* see so-and-so', '*I* hear so-and-so', '*I* try to lift my arm', '*I* think it will rain', '*I* have toothache'. One can point to the difference between these two categories by saying: The cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of an error, or as I should rather put it: The possibility of an error has been provided for. … On the other hand, there is no question of recognizing a person when I say I have toothache. To ask 'are you sure that it's you who have pains?' would be nonsensical. Now, when in this case no error is possible, it is because the move which we might be inclined to think of as an error, a 'bad move', is no move of the game at all.
>
> Wittgenstein 1958: 66–67

The linguistic feature that Wittgenstein describes in this passage also has a home in the mental realm. We already encountered the two different uses in a different disguise when we distinguished between thinking *de dicto* and thinking *de se*. In the former case the thinking subject identifies her intentional object via some description that's supposedly satisfied. And then, the subject ascribes some property to that thing. In the case of Valentina Tereshkova she's identifying her intentional object via the description 'the first woman in space' and subsequently ascribes to that identified woman the property of just having turned 80. However, in the *de se* case the subject doesn't need to do anything like that. Her thought can in a way be directly about herself without any intermediate 'picking-out' to be done. She just ascribes the property to herself without first needing to identify her intentional object in some way or other.

However, the scope of the distinction that Wittgenstein has in mind isn't about the difference between *de dicto* and *de se* thinking. It's much narrower and more specific. We can distinguish the way the intentional object is identified even within different kinds of *de se* attitudes. If we

take the examples from the quoted passage as a guide to understanding the point that Wittgenstein was intending to make—something which is best left to Wittgenstein experts, a set which I'm not a member of—it certainly looks like the former kind are about aspects of the subject that are accessible to anybody, such as her bodily properties. Whether my arm's broken can be determined by a doctor just as well as me—or even better. Whether I've grown six inches is normally determined by someone other than me using a ruler—though I could do it too. It seems that the 'uses as object' apply to properties of subjects which aren't necessarily first-personal but can be ascribed from the third person just as well—for instance our bodily properties.

What other properties of subjects are there which are more readily characterised as first-personal? Classically, mental properties such as our experiences, our beliefs, or our tryings are thought to be intrinsically first-personal. It seems that *my trying* to lift my arm isn't something that can be determined by others—not even using an fMRI scanner. Equally, *my experiencing* a toothache seems to be only accessible to me. How could anybody but me myself know what I'm currently experiencing? It seems that we're now confronted with properties of subjects which are typically first-personal (A.1.5). If we prescind from science fiction scenarios for the moment, we can say that we're typically in a position to know who we're thinking about when thinking *I'm in pain*. As Wittgenstein observes, there's a distinctive oddness in asking whether the subject is sure that it's *herself* who's in pain in such cases. Who else would she be thinking about?

We could take this as a hint that our method of knowing who we're thinking about in such cases is in a way *direct*. In other words, we don't need to first 'pick out' an object in the world before we can ascribe toothache to that object. We just ascribe that property to ourselves directly; no big deal. But this doesn't answer the question why the specific identificational error—namely, misidentifying the object that brought about our thought—isn't possible in the case of ascribing pain to oneself in the *de se* way.

One, possibly better, answer is: being in pain is a kind of first-personal mental state that's usually grasped through introspection, our ability to think about our own mental going-ons. And in introspection the question of *who* we're thinking about doesn't arise since it's not subject to any kind of doubt—something we've learned from Descartes's

famous *Cogito* argument in Meditation II of his *Meditations on First Philosophy*. In this most famous piece of philosophy Descartes argues that some of our thoughts which are about ourselves imply the knowledge of the existence of the thinking subject:

> But I convinced myself that there was nothing at all in the world, no sky, no earth, no minds, no bodies. Did I therefore not also convince myself that I did not exist either? No: certainly I did exist, if I convinced myself of something. But there is some deceiver or other, supremely powerful and cunning, who is deliberately deceiving me all the time. Beyond doubt then, I also exist, if he is deceiving me; and he can deceive me all he likes, but he will never bring it about that I should be nothing as long as I think I am something. So that, having weighed all these considerations sufficiently and more than sufficiently, I can finally decide that this proposition, 'I am, I exist', whenever it is uttered by me, or conceived in the mind, is necessarily true.
>
> Descartes 2008: 18

From the fact that I am introspectively aware of my own thinking, doubting, wondering, it follows that there's someone who exists. After all, without someone who's doing all that thinking there wouldn't be any thinking going on at all. This piece of knowledge is beyond any doubt and can thus be taken as a secure epistemic foundation on which to build. Now, the nature of introspection is such that it necessarily informs us about our *own* mental states. It simply isn't possible to be introspectively aware of some mental state without thereby actually being in that mental state.

Even if we're telepathically engaged in some form of proper mind reading, we would still thereby go through the same mental states as the person we're telepathically hooked up to. We would feel what the other person feels, we would think what the other person thinks. So, your introspective awareness of doubting going on in the *Cogito* doesn't just imply the existence of someone but of you *yourself*. Hence, the *Cogito* can be used to reach an important conclusion: In introspection, there's no room for doubt concerning the intentional object of the subject's *de se* thought (A.1.6).

How does this relate to what was said earlier? The subject using the Cartesian method of doubt is employing Wittgenstein's 'use as subject'. And since there isn't any room for doubt concerning the intentional object, the corresponding introspective belief isn't subject to the identificational failure that we found in other kinds of thought. Hence, the corresponding thoughts are prevented from any possible misidentification of the intentional object. Introspection doesn't involve identification because it's not supposed to inform us about anything but ourselves. Accordingly, we might argue that the use as subject in introspective thinking ensures that we're thinking about ourselves. Thinking about someone else in introspection is 'no move of the game at all'—such an activity wouldn't properly be labelled 'introspection'. Therefore, these mental states can't fail to be about us: they're *immune to error through misidentification*.

This peculiar term has been coined by Sydney Shoemaker in his seminal paper 'Self-Reference and Self-Awareness' (1968). To reach a characterisation of the term Shoemaker takes Wittgenstein's catalog of examples very much at face value. He argues that the distinction between the 'use as subject' and the 'use as object' of the first-person pronoun is essentially made in terms of what kind of properties we want to ascribe to ourselves. If they're the kind which are normally grasped through introspection, then the resulting thought is a candidate for immunity to error through misidentification because there's 'no question of recognising a person'—as Wittgenstein would put it—in introspection. Introspection is always about ourselves. And this is why the subject can't wonder who's having a toothache when she thinks *I have a toothache*.

Unfortunately, this simple picking and choosing from properties doesn't work. We can't just distinguish between 'introspective' properties and others and claim that self-ascribing the former results in thoughts that exhibit immunity because it's impossible to introspectively think about someone else. Here's why: Sometimes a subject self-ascribes an introspective property on a strange epistemic basis. For instance, I see that the person in the mirror is writhing in pain and conclude that therefore I'm in pain—but strangely without feeling it. This is because I sensibly believe that it's my mirror image. Unbeknownst to me, this belief is false: the image is of my evil twin. In such a case I'm therefore misidentifying the subject in believing that I'm in pain.

The basis of my belief is the reflection of my twin which I mistakenly take to be of me. And thus, one and the same property leads to a belief with immunity in one case and a belief without immunity in another.

It's true that we usually grasp and self-ascribe introspective properties like *being in pain*, *seeing something*, or *trying to move a body part* through introspection. And in such cases there's in fact 'no question of recognising a person' because the nature of introspection is such that it necessarily informs us about states of the thinking subject. But we can ascribe introspective *properties* in all kinds of contexts; some of which don't have the peculiar nature of introspection. So, Alpha can ascribe the property of *seeing something* on a nonintrospective basis too—for instance when she believes that her friend can't see the beautiful bird. And in such cases there's a question of who the subject's thinking about and thus there's the possibility of misidentification. Therefore, so the objection concludes, the fact that the property is of an introspective kind can't play the demarcating role between thoughts that exhibit immunity and thoughts that don't.

So, we have to rethink our model. We want to know which feature makes some of our thoughts immune to error through misidentification and some not. We saw that Shoemaker interpreted Wittgenstein's distinction between the 'use as subject' and 'use as object' in a way which distinguishes between mental properties on the one hand and bodily or physical properties on the other. His idea was: when a subject self-ascribes mental properties, we're left with a *de se* thought that exhibits immunity. But, as our arguments showed, the simple distinction between two types of properties that we can self-ascribe doesn't suffice. Subjects can ascribe mental properties to themselves in different contexts on distinct and potentially unreliable epistemic bases. And only in some of these cases do we have a candidate for immunity.

A more reasonable way to understand the phenomenon of immunity is to make a small amendment to the original suggestion: Any specific thought—independently of whether it's *de se* or not—is immune to error through misidentification if it's formed on the right *epistemic basis* that doesn't allow for misidentification. But what could this right epistemic basis amount to? Let's take as an illustrative example Beta's belief that she's in pain:

(2) *I'm in pain*

On first sight it seems that Beta can't fail to think about herself in believing (2). We saw that the thinking subject is always the intentional object of her *de se* thoughts. However, this doesn't imply that we're in fact always also basing these *de se* intentional states on a basis that necessarily or actually concerns us. Sometimes we're just plainly mistaken about who we're actually thinking about—even in the cases where we're consciously thinking about ourselves. In the case of *de dicto* thinking we've already encountered the possibility of misidentification. What Beta takes to be the tree in her garden might as well be her friend. In that case the intentional object of her thought doesn't correspond to the thing which Beta takes as the basis of her thought. So, we can be wrong about what we're actually thinking about. But how's that possible in *de se* thinking? Isn't that always about us?

Let's contrast two scenarios in which Beta could come to believe (2). In the first scenario, she forms the belief on the basis of her own proprioception and nociception. Because she feels the burning pain in her calves from running earlier she subsequently believes (2). In a case where everything works as it normally does, she receives proprioceptive and nociceptive information from her own limbs in the regular way. And forming the belief *I'm in pain* on that specific epistemic basis is a way that ensures that the subject can't misidentify who she's thinking about. There's no wedge that could be driven into the epistemic process underlying the formation of (2) on the basis of a subject's first-personal experience of pain. Hence, there's no possibility of misidentifying the subject of pain if the relevant belief is formed on that appropriate basis. Our proprioception normally informs us about our own states. So, *pace* science fiction scenarios, attitudes based on this form of perception necessarily concern us. And *ergo*, the belief (2) is immune to error through misidentification in this scenario.

Now, let's look at a second possible scenario in which Beta can come to believe (2). Here we have to make use of some science fiction but that only makes it that much more interesting. In this scenario, Beta has been using strong pain killers which effectively remove all feeling of pain, were there any to appear. Now, Beta is examined by a doctor using some elaborate machine that detects painful injuries. She informs Beta on the basis of these scans that her calves are heavily inflamed—a usually very painful injury. Beta then reacts to this piece of information by saying something like 'Oh my, that means that I'm

in pain'. And this again is a good indicator that she now believes (2). However, unbeknownst to either of them, the doctor made a mistake and inspected the scan of someone else—Gamma. So, Beta forms the belief (2) on a different epistemic basis than in the earlier scenario. Proprioception and nociception, which in the regular case give rise to beliefs that exhibit immunity, didn't play a role this time. Instead, Beta forms her belief on the basis of the doctor's third-personal testimony.

Does this lead to the possibility of misidentification in her belief (2)? Yes. The fact that it's possible for the doctor to misidentify the origin of the scans opens the possibility for misidentification for any thought that's based on her testimony. The epistemic 'impurity' seeps through. It's in virtue of the doctor's *de re* belief *Beta is in pain* being subject to identificational error together with the fact that Beta's *de se* belief is based on the doctor's belief that Beta's *de se* belief is now also error-prone. We can reconstruct Beta's reasoning process in the first scenario in the following way. She has a certain first-personal painful experience. These experiences are such that they usually give rise to corresponding attitudes which are immune to error through misidentification because the things we feel are normally about us. Subsequently, Beta forms the *de se* belief (2) on the basis of this first-personal experience. Hence, her reasoning is secured against error through misidentification; we normally can't experience someone else's feelings. In contrast, Beta's reasoning in the second scenario begins with the belief that the doctor's information is about herself—and that belief isn't immune to misidentification. The information could be about someone else. So, even if the belief (2) is classically formed on a secure basis, it's not necessary that it's formed on such a basis—there are deviant cases where we can go wrong. What's the upshot of this argument?

Immunity to error through misidentification doesn't have its source in the nature of the ascribed property but in the nature of the epistemic basis on which the relevant *de se* belief is formed. This implies that not all *de se* beliefs exhibit immunity. Beta's belief (2) in the second scenario is still *de se*, she's believing something about herself in the *de se* way. However, due to the unreliable nature of the epistemic basis, the question 'Are you certain that it's you that's in pain?' isn't nonsensical anymore—it's now a legitimate move of the game.

We owe this crucial insight to Gareth Evans (1982). His precise position is extremely complex and somewhat cumbersome. Hence, I

won't elaborate it here in detail. However, we'll make use of Evans's notion of *identification-freedom* in order to explain why some of our thoughts are subject to misidentification and some aren't. Evans's idea is that all thoughts which are identification-free are *ipso facto* immune to error through misidentification. The obvious question is: In which cases is a subject's thought identification-free? To answer this question we can start by looking at our ability to think about things in general. Here, we find clear cases of identification that can be contrasted. In the most simple case a subject picks out an object in the world and ascribes a property to it. For instance, in believing *Sydney is the capital of Australia* Alpha picks out a famous city and ascribes to that thing the property of *being the capital of Australia*.

But we have to be careful here. This way of describing the case doesn't imply that the subject literally *does* two things in believing. It's rather that we can analyse this belief by distinguishing two elements that are present in the thought. First, there's an ascription element of the form *x is F* which ascribes a specific property to a thing. Secondly, there's an identification element such as *Sydney is the x* which identifies the thing the subject's thinking about. Most of our thoughts comprise these two elements. For instance, when Valentina Tereshkova believes that the first woman in space just turned 80 she identifies a thing in the world (*the first woman in space is the x*) and then ascribes the property of just having turned 80 to that thing (*x just turned 80*). Similarly, in believing that the shop is open we identify the shop we're thinking about—it's the one at the corner—and ascribe to that specific identified thing the property of *being open*. Hence, whether we believe in the *de re* or *de dicto* way, as soon as the subject has to identify the thing she's ascribing a property to in thinking we have a case of an attitude that involves an identification element.

Now, let's look at thoughts which are immune to error through misidentification against this background. One rather obvious way of accounting for this feature is by using their supposed identification-freedom as an explanation for immunity: They don't involve an identification element and hence the intentional object can't be misidentified. The property is simply ascribed to the appropriate thing. So, when Beta believes (2) on the basis of her proprioception she simply ascribes pain to herself—the appropriate intentional object of beliefs of that kind. Whenever the question 'Who's this thought about?' is

superfluous we're confronted with a thought that's identification-free. Of course, this opens up a whole new can of worms. How is it possible that some thoughts are identification-free and others aren't? How is it possible that one and the same thought—such as our belief (2)—can be identification-free in one case and not in others? A theory of *de se* thought is supposed to give answers to these questions and I'll provide them in this book.

The idea that identification-freedom, and not some special class of properties, is responsible for a thought's immunity has an interesting consequence: we're capable of self-ascribing all kinds of properties and the resulting thoughts might be immune to error through misidentification. As long as the *de se* thought is identification-free we're on the safe side. This is in stark contrast to Shoemaker's account. He focused his explanation of immunity on a specific kind of property which we ascribe—introspective ones—whereas Evans's account is not restricted in this way; here the epistemic basis is what's relevant. We can self-ascribe mental properties just as well as bodily properties. Beta can believe *I'm thirsty* just as well as *I've got ink on my chin*. And as long as the thought was formed on the right epistemic basis it can be immune to error through misidentification. What would be an example of such an immune belief which consists of the self-ascription of a bodily property? Here's an example:

(3) *My legs are crossed.*

Using Evans's account, we can easily explain why (3) might exhibit immunity. On the one hand, Beta might see her reflection in the mirror. She sees that the person in the mirror has crossed legs. She then identifies herself with that person and ascribes to herself—via this identification—the property of having crossed legs. So, her belief involves an ascription element as well as an identification element. Accordingly, this way of forming the belief—via a visual experience of one's own body and the identification of that body as one's own—doesn't give rise to immunity. It involves an identification which can be erroneous because the source of the visual experience could be someone else. Another example where a *de se* belief fails to exhibit immunity is when a subject believes (3) on the basis of seeing a picture of herself. Again, the belief is based on the fallible recognition of herself in the picture. And thus, it isn't immune to error through misidentification.

On the other hand, if Beta thinks (3) on the basis of her proprioceptive feeling she needn't identify a thing in the world as herself before she can ascribe the property to herself. In the contrasting mirror case, Beta's belief relies on her holding something like the belief *That woman is me*. Her belief (3) thus depends on some identificational premise. And this epistemic dependence—her justification for believing (3) is dependent on the mentioned identificational belief—results in the possibility of misidentification in her believing (3). However, such a dependence is absent from the case where Beta forms the belief (3) on the regular basis of her proprioceptive feeling. She feels that her legs are crossed and directly self-ascribes the property of having crossed legs. In this case, her belief doesn't get its justification from some identificational premise that could go wrong. When we hold beliefs on the basis of proprioception we don't need to identify the intentional object of our mental attitude: no identification is involved and thus the belief is immune to error through misidentification.

This shows that *de se* thoughts involving introspective properties like *being in pain* or *seeing something* don't have a monopoly on immunity. Other properties are on the market too. Furthermore, it shows that one and the same thought can be formed on different epistemic bases. And depending on the nature of that basis the resulting thought can be properly 'vaccinated' or not. Evans's idea of identification-freedom neatly explains how this difference arises. Immunity occurs in all those cases in which a belief or judgement doesn't involve an identification—neither as an explicit element nor as part of the epistemic warrant.

1.3 KNOW THYSELF

Now that we've discussed this first epistemic feature of *de se* thoughts we can dedicate ourselves to the historically most important epistemological aspect of our thoughts about ourselves. We're now talking about the profound aspect of *de se* thinking that's referred to in the inscription on the Temple of Apollo at Delphi: *know thyself*. Self-knowledge is oftentimes thought to be special because it's more secure and reliable than regular knowledge. It originates from different sources and proceeds along distinct epistemic 'paths' compared to our acquisition of knowledge about the external world. It's not surprising that Descartes's rock bottom piece of knowledge results from self-knowledge: the in-

dubitableness of *I think, I am*. But what's special about self-knowledge so that it's provided these extraordinary epistemic qualities?

Let's look at one classical instance of self-knowledge that's referred to in Plato's 'Apology'. In that dialogue Socrates hears of the Delphic oracle claiming that no one in Athens was wiser than Socrates himself. This puzzles Socrates since he didn't believe himself to know many things about the external world:

> Finally I went to the craftsmen, for I was conscious of knowing practically nothing, and I knew that I would find that they had knowledge of many fine things. In this I was not mistaken; they knew things I did not know, and to that extent they were wiser than I. But, men of Athens, the good craftsmen seemed to me to have the same fault as the poets: each of them, because of his success at his craft, thought himself very wise in other most important pursuits, and this error of theirs overshadowed the wisdom they had, so that I asked myself, on behalf of the oracle, whether I should prefer to be as I am, with neither their wisdom nor their ignorance, or to have both. The answer I gave myself and the oracle was that it was to my advantage to be as I am.
>
> Plato 1997: 22, 22d–e

Socrates's knowledge amounts to knowing something that's distinct from knowledge of the external world and its workings. It's knowledge of the fact that he himself didn't know many things—least of all *all* things. It's the archetype of realising the limits of one's own knowledge. He didn't know how to build a ship or write a beautiful poem, but he knew about his ignorance in this respect. In contrast, the craftspeople thought that their expertise was boundless and that they knew all there was to know. But, Socrates wouldn't have been able to attain this insight without the capacity to think about himself while grasping that he's doing so. So, Socrates's knowledge is crucially based on his ability to have *de se* thoughts.

We can now see how the dependence of self-knowledge on our ability to think about ourselves is crucial for our investigation of *de se* thinking. All items of self-knowledge are instances of *de se* thinking because

they appear in the first-person form. Hence, without *de se* thinking you wouldn't be capable of gaining self-knowledge because you wouldn't realise that whoever you're thinking about is actually you. When thinking about how *de se* thinking works we therefore need to make room for the possibility that some of these instances of thinking amount to self-knowledge with all its distinctive features. The idea being that the transcendence of self-knowledge originates in our unique ability to think in the *de se* way.

So let's look at the nature of self-knowledge and what elevates it above other kinds of knowledge. First and foremost, what applies to all things in life must also apply to knowledge: you win some, you lose some. Possible items of self-knowledge share this characteristic with other possible items of knowledge. So, when we're interested in the question of self-knowledge that means that we sometimes get it wrong when we self-ascribe a property. Supposed self-knowledge is not beyond doubt. For instance our legs weren't crossed after all, so our claim of self-knowledge on the basis of *de se* believing (3) is unjustified. At other times, we're getting it right: Beta is really in pain and really wants some pain killer. Everything in her proprioceptive and nociceptive system works as it should be and she's not forming her *de se* belief in some deviant way. Hence, she's justified in believing (2) and has a valid claim to self-knowledge in this case.

The goal now isn't to give an account of self-knowledge but rather to identify the general nature of the phenomenon and relate it to our ability to think in the *de se* way. Let's start with a very plastic subtype of self-knowledge. It concerns our knowledge of our own mental states, more specifically our beliefs. It's in these cases where we can find some of the most characteristic features of self-knowledge. To take an example, I believe:

(4) *Roger Federer has so far won 20 Grand Slam singles titles.*

Furthermore, I know that I have this specific belief—whether it's correct or not. Now, someone might come along and ask me two different kinds of questions: 'Do you believe (4)?' and 'Why do you believe (4)?' The answer to the latter question might involve aspects that have nothing in particular to do with myself. I'm giving an explanation of why I came to believe what I believe by pointing to the public sources of this belief. For instance, I could answer that I've read the relevant

Wikipedia article, that I've heard this claim uttered by many people, or even that my neurons are connected in such a way as to cause in me the relevant belief. It's true that the last option is less likely because we normally don't explain why we came to certain beliefs by pointing to our brains. Regardless, it shares with the other answers the characteristic that I don't have any kind of special position regarding the knowledge in question, i.e. the sources of my belief. The reasons why I believe what I believe are public. They're not in my head. You could go to Wikipedia and check. Or you could follow me around and listen in on all my conversations. The reasons why I'm in a better epistemic position are entirely contingent on the fact that it's easier to know what *my* evidence is. But you could gather all the same evidence and thus the epistemic discrepancy could be levelled.

On the other hand, I seem to be in a special position to answer the first question. More specifically, I'm the touchstone of whether or not I have a specific belief—manifesting my authority relative to the item of knowledge—and I'm in a very good epistemic 'position' to determine whether I believe something or not. Gianfranco Soldati puts this point in the following way:

> [Self-knowledge] is a domain where *epistemic authority* of the subject is particularly manifest. I know what I think without having to rely on the kind of evidence you would typically have to rely on in order to know what I think.
>
> Soldati 2013: 169

Soldati thus identifies two characteristics of self-knowledge. With regard to our own beliefs we're generally *epistemically authoritative* and *epistemically well positioned*. First, we're usually in a situation where my word trumps yours. If I claim to believe (4) and you deny it, my claim normally overrides yours. I have authority over the possible items of self-knowledge concerning me. So, the idea that we're epistemically authoritative reflects the oddness of being told by someone else that—according to her analysis of my behaviour, the oxygen flow in my brain, my dreams, or what not—I don't *really* believe (4) at all. In a way this is similar to the oddness of being told by someone else that you don't actually prefer tomatoes over eggplants despite feeling that way.

What's the source of this oddness? Usually our first-person perspective informs us pretty well about what we believe, prefer, or desire. And

this results in my astonishment were you to tell me that I don't actually believe (4) despite the fact that introspection tells me the opposite. How would you know better than me what I believe or not? Imagine you ask me how many titles Roger Federer has won. I'll answer '20 and he might even win a 21st title before retiring'. If you were then to tell me 'Yes, I'm aware of the fact that it seems to you that you believe (4). But listen, you're actually wrong about that. You don't really believe that!' I would strongly believe that you're messing with me and not really questioning that I believe (4). We're better at telling what we believe than some neuroscientist's fMRI scanner. This is what we mean by being authoritative regarding knowledge of our own mental states. Not only am I in a good epistemic position to know what I believe but 'my word is truth'.

Secondly, I usually come to knowledge about my beliefs on a different epistemic 'path' than someone else who wants to know what I believe. If you want to find out about my beliefs, you've got several options: you could ask me directly, you could observe what I do and try to infer what I believe from it, or you could put me in an fMRI scanner and do some complex, and so far unreliable, procedure in order to 'read off' what I think from the provided data. In all these cases, you have to arrive at knowledge about my mental states in some *indirect* way. I, however, don't have to do anything of that kind. Of course, I could do these things as well—and with some of my mental states it might even seem necessary to see a psychoanalyst—but it usually isn't necessary. I have a more direct and immediate access to what I believe or not: I don't have to ask, observe, infer, or 'read off'. I just know.

There are different ways to spell out what this direct knowledge amounts to and where it originates. One option is that the relation we're in to ourselves is of a direct and ontologically primitive nature. The idea is that every subject is immediately acquainted with herself and thus epistemically privileged regarding herself. There's no better epistemic point of view on a subject than actually *being it*. A second option is that introspection—the capacity to be acquainted with one's own mental states—offers us a qualitatively exquisite insight into our own minds. Consequently, we would be in a special position with regard to self-knowledge because introspection acquaints us directly with the states in question whereas others have to rely on their indirect perception of us in order to gain knowledge of our minds.

According to both options there's a distinct asymmetry between knowing one's own mind and knowing another's mind. This asymmetry is grounded in the fact that we have direct knowledge of our own mental states and only mediated knowledge of those of others. This reflects the proposed fact that subjects are epistemically well positioned regarding knowledge of their own mental states.

There might be an intimate connection between the fact that we sometimes are in a good position to know our own mental states and our authority regarding that item of knowledge. Why is that? Imagine you witness a car crash. Later that day you hear a contradicting report from someone who heard about the same car crash on the radio. While you believe that the red car drove into the blue car, the other person claims the opposite. In this case, it seems that you're authoritative about the details of the crash *because* you were in a better position to gain knowledge about the crash than a person who relies on mere hearsay. In the same way we can take Soldati as claiming that you're epistemically authoritative with regard to items of self-knowledge *because* you usually don't have to rely on the kind of evidence that's typical of third-personal knowledge. You're in a privileged position and this provides you with the authority in question (A.1.7).

To be clear, not all possible items of self-knowledge have these two features. There are many instances of *de se* beliefs which are utterly mistaken. For instance, Beta might think that she's a charitable person while always failing to donate to charitable causes. But if Beta doesn't actually exhibit any charitable behaviour, we are justified in denying her privileged position regarding knowledge of her own character traits. She's a bad judge of her own character. And were Beta to claim that she's really a charitable person, it would be ok to agree to disagree. Beta's self-deception, which isn't based on any reasonable justification but on something else, undermines her authority with regard to purported self-knowledge. Our authority isn't global because our capacity for self-knowledge isn't infallible (A.1.8).

Character traits aren't the only possible items of self-knowledge where subjects oftentimes lack both epistemic authority and a privileged access. Other examples include our own moods and emotions. I might think that I'm in a good mood while I'm actually constantly shouting at people and criticising them—something rather atypical of being in a good mood. Furthermore, we seem to be rather weak in

figuring out the causes of our own actions. Beta might think that she bought a new smartphone because it's superior to the previous model which she claims was already having some serious issues with the camera, generally sluggish, and so on and so forth. But really, she just wanted to have something new, maybe as a status symbol, or because she enjoys new gadgets. This kind of epistemic impairment also concerns assessments of how we would feel were certain things to happen in the future. For instance, Gamma might think that having a child will make her happier, but in fact the opposite might be true.

Looking at self-knowledge from this empirical point of view thus exposes some very central and wide-ranging defects and shortcomings concerning the scope and content of self-knowledge. There might be cases where we're both privileged and authoritative, but the exemptions are aplenty. However, the original claim wasn't that we're *always* epistemically well positioned and authoritative. Rather, the idea is that we're *usually* or characterstically in this situation. In a manner of speaking, the burden of proof lies with someone who wants to deny our authority and privileged access. It's the others who have to demonstrate that we're a bad judge of our own character traits, emotions, or moods, and that in these specific cases our word isn't truth. And the discussed shortcomings merely demonstrated that there are certain limits to our authority and epistemic position which have to be fathomed.

There are some, however, who outrightly deny even the *possibility* of self-knowledge. Wittgenstein features most prominently within this camp. However, he doesn't raise empirical objections but rather conceptual ones: self-knowledge isn't knowledge at all. To support this rather surprising claim he argues in his *Philosophical Investigations* (1953) that supposed knowledge of our own mental states isn't an *epistemic achievement* and thus can't take credit for being successful—something characteristic of knowledge. It's an achievement to be right about what someone else is believing, desiring, or feeling. We employ a certain method of acquiring knowledge of others' mental states or facts about the world in general. We observe, infer, or conclude from certain information that things are thus and so. Similarly, it's an achievement to calculate the solution to a complex mathematical problem; there's a significant risk of being wrong. But in the case of self-knowledge, there's no similar epistemic *method* of acquisition:

> I can know what someone else is thinking, not what I am thinking. It is correct to say 'I know what you are thinking', and wrong to say 'I know what I am thinking'. (A whole cloud of philosophy condensed into a drop of grammar.)
>
> Wittgenstein 1953: 222e

Again, I'm not claiming expertise in Wittgenstein exegesis, but we can take home two points from this quote and the surrounding passages. Both concern the desideratum that for something to qualify as knowledge it needs to be an epistemic achievement. And because it needs to be achieved it has to satisfy two conditions. First, possible items of knowledge should be capable of being false. Where there's no possibility of error there's no possibility of knowledge. Secondly, possible items of knowledge need to be susceptible to public verification—a central theme of Wittgenstein's philosophy. It seems necessary that we can check whether someone's claim to knowledge is fulfilled or not. If there were no way for you to tell whether I really believe (4), then it wouldn't make sense to call my claim to know what I believe a possible item of knowledge because there wouldn't be a way for others to challenge my claim.

Wittgenstein's argument against the possibility of verification of first-personal mental states is connected to his well-known and complicated *Private Language Argument*. In its essence it's meant to show that saying of something that it has a specific property entails that there is some way for us to distinguish the cases where it has that property from the cases where it doesn't. In other words, when we're concerned with knowledge, 'anything goes' isn't a viable option. And that means that we need some kind of publicly intelligible criterion to distinguish the have's from that have not's. Without some independent standard of evaluation one's claim to knowledge can't be verified. The crucial question is then: What are the independent standards—the criteria—on the basis of which we can determine whether my claim of knowing that I believe (4) is true or not?

As a response to this question Wittgenstein sets up a dilemma for anyone trying to establish some extraordinary form of self-knowledge that's based on some privileged first-personal access. Either the criteria on the basis of which I self-ascribe a property are of a public nature or they are of a private kind. The latter possibility would lend support

to the idea that there's something special about self-knowledge which gives rise to its purported authority, while the former possibility undermines this relationship. So, our proponent of a special kind of self-knowledge need to opt for the latter position. But the problem with the idea of private criteria of correctness runs against our commitment to knowledge as something that's intersubjectively determined. My claims for knowledge can be challenged and supported on the basis of public information. Were my claims of self-knowledge based on purely first-personal data, such a challenge would be pointless. How could I determine whether my self-ascription is correct without some independent and graspable criterion? As Wittgenstein claims: 'One would like to say: whatever is going to seem right to me is right. And this only means that here we can't talk about "right"' (Wittgenstein 1953: §258).

The upshot of Wittgenstein's dilemma is thus the following: The criteria for genuine self-knowledge—which is supposedly distinct from and paramount to other kinds of knowledge—are either private, but then they're null and void, or they're public. And if they're public, then there's nothing special about self-knowledge since it's just another form of regular knowledge about things. It collapses together with the undermined pillars of authority and privileged access (A.1.9).

Wittgenstein's criticism shouldn't discourage the project of explaining self-knowledge. After all, it doesn't explain the profound oddity of having one's supposed items of self-knowledge challenged. Thus, the empirical and conceptual doubts about the distinctive nature of self-knowledge can be acknowledged but not taken as conclusive reasons to abandon the possibility of self-knowledge altogether.

Furthermore, there are several concrete options to respond to the challenges we encountered. One option is to remain mute and neutral about the possibility of self-knowledge. In effect, we can take an agnostic point of view on the debate. Since it's far from clear whether one of the two sides clearly has the better arguments, this option simply acknowledges the controversial nature of the topic and tries to work its theory of *de se* thinking around that fact. Another option is to claim that we can still uphold our privileged access to our mental states as the distinctive mark of self-knowledge—while accepting that knowledge has some public element. In effect, it rests content with the claim that there are different kinds of knowledge—knowledge by acquaintance,

knowledge by testimony, or also knowledge by privileged access. Self-knowledge is just one of these special kinds of knowledge.

For the purpose of explaining the nature of *de se* thinking we'll employ a combination of these two strategies. One the one hand, we only have to account for the possibility of self-knowledge without settling in on any substantial theory. In other words, we only need to be capable of fitting into our account the possibility of self-knowledge—just in case that it proves to be a real and distinctive thing. On the other hand, some self-ascriptions of properties are indeed special. And this might amount to something like self-knowledge based on privileged access. However, should this claim prove unjustified, we can still fall back on the first strategy.

1.4 WHO'S MAKING THAT MESS?

The last feature of *de se* thinking that we'll examine is the connection it has to our actions and behaviour. Here's the reasoning behind this conceptual link: Many of the things we do, we do for a reason—normally something that motivates us to do something or not. But we don't just act for any reason, it has to be a reason *for us*. And taking a reason as being a reason for us is to exert one's capacity for *de se* thinking.

To take an example, if Valentina Tereshkova believes that the first woman in space needs to buy groceries, she's not necessarily motivated to go shopping herself. That's because she doesn't need to know that she herself is the first woman in space in virtue of her *de dicto* belief and she maybe couldn't care less what that woman needs to do. On the other hand, if I believe that I'm in pain, I'm *ceteris paribus* very motivated to do something about it. I'll go to the medicine cabinet and get some pain killer or try to avoid what's causing me pain in some way. Were I to entertain a mere *de re* or *de dicto* belief about myself, I wouldn't immediately be motivated to do something about it. Why's that so?

John Perry provides us in his 'The Problem of the Essential Indexical' (1979) with a nice thought experiment that illustrates why *de se* thinking embodies the motivational power that's needed for action. It's also meant to demonstrate why other beliefs lack this important motivational power:

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

The story manifests a specific type of cognitive move: from one kind of belief to another. First, Perry believes *de dicto* of some person or other that she's making a mess. And after a while he acquires a new belief and now believes *de se* that he himself is making a mess. This new belief is of special practical and motivational importance to him. It's only because he finally believes *de se* about himself that he's making a mess that he's now driven to rearrange the torn sugar sack in his basket. At first, Perry didn't feel motivated to clean up the mess because he didn't believe *he himself*, i.e. in the *de se* way, was making a mess. Only once he entertains the relevant *de se* belief

(5) *I'm making a mess*

is he motivated to act accordingly. And if he entertains any non-*de se* belief about himself without a supporting *de se* belief, he won't contemplate that he's the potential mess-maker.

Can we give some justification to this claim? As so often, mirrors will help us to get the main point across. If he sees a reflection of himself in the mirror making a mess without realising that it's his own reflection, he's *de dicto* thinking about himself without believing (5). By seeing the mess-maker in the mirror he might think *That person is making a mess*. And because he doesn't recognise himself in the mirror, he won't do anything about the mess he's making. And the case is similar for any potential *de re* belief about himself. Believing *John Perry is making a mess* wouldn't automatically move him to rearrange the torn sack. This is because he might have forgotten his own name or how he looks in the mirror or be subject to all kinds of weird science fiction scenarios. As long as he doesn't think that he *himself* is making a mess by believing something along the lines of (5), he won't stop and inspect his basket.

These examples lead to the idea that *de se* thinking is intimately connected to our behaviour and motivation for action. It's only in virtue of entertaining a reason in the *de se* way that a subject feels moved to act accordingly—and, if circumstances allow it, will act as well. The argument for this connection runs by showing that everything that motivates a subject to act contains at least one *de se* element. This claim isn't uncontested and I will now provide some justification why we should nonetheless hold on to the idea.

We'll use Perry's example as an illustrative case study. The first step of the argument is to give all the possible ways that Perry could think that he's making a mess. The *de se* belief (5) is an obvious first example, but due to the fact that it's our standard *de se* belief, it's also the last resort for our opponent who's trying to show that a *de se* belief isn't necessary for action. After all, she doesn't want *de se* thinking to be involved.

What other options are there? The obvious candidates are *de re* and *de dicto* thoughts about himself. We already saw that *de re* believing *John Perry is making a mess* won't motivate him to clean up the mess. In the ultra-amnesiac case, he doesn't care what John Perry's doing. He needs to identify himself with this person in order to be motivated by what John Perry's doing. So, only by thinking something like *I'm John Perry* will the *de re* belief in question give him a reason to clean up the mess. But this new thought is a classical example of a *de se* belief. Hence, a mere *de re* belief isn't enough to motivate a subject to act because only in conjunction with a *de se* attitude does it provide a reason for the subject to act.

What about the *de dicto* option? To make our case as strong as possible we can use the most detailed way of describing John Perry. The description in question is so detailed that it only fits this specific person in this specific possible world. It's so detailed that the smallest deviation will give rise to a new distinct description. Let's call this detailed description α. Now, would Perry be motivated to clean up the mess he's making were he to believe α *is making a mess*? Because the description is so specific he can't fail to refer to and think about a particular person, namely himself. But all that applies to any kind of detailed description. He also can't fail to think about a very specific person were he to think β *is making a mess* and where β is the ultimately detailed description that uniquely fits Roger Federer. The fact

that he can't fail to think about himself on the basis of the description α is irrelevant to the question whether it's enough to motivate him to act. Because the messes that Roger Federer is making aren't a concern of his—despite the fact that he can't fail to refer to Roger Federer using the description β—what α is doing isn't either. Furthermore, why should this very specific thought involving the description α only motivate Perry to act and not other subjects who also think about Perry using that description? The point is that merely knowing of an incredibly detailed description of himself doesn't yet ensure that he's aware of the description referring to himself. So, only in conjunction with the *de se* belief *I'm* α will this incredibly detailed *de dicto* belief motivate him to do what it would motivate the referent of α to do. The case of *de dicto* belief hence doesn't help us either. As in the case of *de re* thinking, we need the addition of a *de se* attitude to generate a reason for a subject to act in a certain way.

This already provides some initial support to the idea that all our actions are based on some form of *de se* thinking. We've shown that merely entertaining a *de re* or *de dicto* state is insufficient to motivate the subject to act accordingly. The suggestion was that the reasons for our actions need to be reasons for us, for otherwise they wouldn't concern us directly and wouldn't bring about action. The fact that we slaughter more than 150 billion sentient animals every year is only insofar a reason for me to change the way I eat if I relate this reason in some way to me. This might be by desiring to stop supporting the industry that causes these deaths. Such a desire, however, necessarily comes in the *de se* form: *I don't want to eat animals anymore*.

An example which takes place on the lower unconscious levels of *de se* thinking is additionally illustrative and supportive. Consider that I'm now sitting in my living room and feeling some thirst. This produces a *de se* mental state of the form *I'm thirsty*. I'm not necessarily consciously entertaining this thought. It's merely there in my mind. Implicitly, I additionally know that in order to quench thirst one needs to go to the kitchen to get water. But my knowledge of this instrumental connection doesn't itself produce the appropriate action. After all, it applies to anyone who's thirsty and has water in the kitchen. Were I not thirsty, there wouldn't be a reason for me to go to the kitchen merely on the basis of knowing of this instrumental connection. Only in conjunction with my *de se* mental state does my knowledge of

the connection between thirst and the water in the kitchen provide a reason *for me* to go to the kitchen.

We can now see more explicitly why explanations of actions make reference to the subject's *de se* thoughts. Both from the first and third person, explanations are only complete if they make implicit or explicit reference to some *de se* thought or other. If we look at Alpha's reasoning 'I ran home because it was dark', we get an explanation of her running home only because it gives us implicit information that Alpha doesn't like being in the dark. A complete first-personal explanation would have the form 'I ran home because it was dark *and I'm afraid of the dark*'. Similarly, we only explain someone's action completely by incorporating a *de se* thought of hers. For instance, when we want to know why Beta watered the plants it's insufficient to say that the soil was dry. We also have to mention that Beta believes that dry soil leads to withering plants. And crucially, we've got to include Beta's desire for living plants which she would express by saying 'I want living plants'. And this, of course, is a report of some *de se* thought she has (A.1.10).

Being motivated to act thus requires some form of *de se* thinking which ties the reasons in the world to our own motives. It's not enough to desire some state of the world in order to change it. One has to intend to do something about it first. And this intention has the form *I will do* φ—a clear case of *de se* thinking. This demonstrates the intimate connection that our ability to think about ourselves has to the explanation and motivation of our behaviour and intentional action.

1.5 BEGINNING FROM THE ORIGIN

We've now established five features that are important to our discussion of *de se* thinking. First, all *de se* mental states are about the thinking subject. In other words, the thinking subject is necessarily the intentional object of the thought. Secondly, the satisfaction conditions of *de se* thoughts depend in a systematic way on the thinking subject. Whenever a subject entertains some *de se* thought she will be part of what determines the conditions of satisfaction of that episode of thinking. Thirdly, some of our *de se* mental states are immune to error through misidentification. We can't misidentify the intentional object of our thoughts—i.e. ourselves—in some special episodes of *de se* thinking. Fourthly, *de se* beliefs provide the fundament of self-know-

ledge. Only because we're capable of thinking about ourselves in the *de se* way are we also able to achieve self-knowledge, by knowing that this knowledge pertains to ourselves. And fifthly, our intentional action and behaviour is intimately connected with our ability to entertain *de se* thoughts. When we explain our action we make reference to *de se* mental states and the reasons that motivate us are reasons we grasp as our own by way of *de se* thinking.

What kind of account of *de se* thinking can do justice to these identified features? Luckily, contemporary philosophy has offered us a selection of interesting options to choose from, so we may not need to pull a new one out of our hats. The goal of the thesis is to determine how well some of these options fare in dealing with the different features we've established. We'll see that some of them are better than others at vindicating a specific feature. But the same ones might be worse off in dealing with other features. In this way, the discussion of our theoretical options won't aim to exclude a specific theory by principle. Rather, I'll identify problematic and felicitous aspects of these strategies with respect to the features of *de se* thinking and determine their merit on the basis of this analysis and their overall coherence and plausibility.

None of the employed strategies that will be discussed is free of problems and obstacles. But some of them are less problematic or better at dealing with the overall picture of *de se* thinking that has been illustrated in this first chapter. The goal of the book is then to provide an alternative story of how to characterise *de se* thinking. This new and more encompassing strategy is supposed to account in the most coherent way all these characteristic features of *de se* thinking.

But what's the overall picture that I'm providing? A detailed account will have to wait until chapter 4, where I'll have the chance to develop the theory in detail against the backdrop of our previous discussion of competing options. But we can end this first chapter with a short synopsis of my favoured view. This synopsis is intended as a rough sketch against which we can compare and contrast the other theoretical options. For this purpose, I'll refrain from giving any arguments and supporting reasons and merely state the different characteristics outright. The arguments will follow later.

The guiding principle of the account is that subjects self-ascribe properties when they entertain *de se* beliefs. For instance, when Alpha believes (3) she self-ascribes the property *having crossed legs*. But,

as we saw, we can self-ascribe properties in different epistemic contexts. In some cases, a self-ascription is derivative and based on some prior ascription. But in other cases, I'll hold that the self-ascription is *primitive*. This primitive self-ascription of properties forms the basis of *de se* thinking. If Alpha experiences the feeling of having crossed legs, she can thereby primitively self-ascribe the relevant property—resulting in the *de se* belief (3). And this makes it the case that she takes *herself* to have crossed legs.

In this way, the account I'm defending is based on the concept of primitive self-ascription. This again can be characterised in more detail using the notions of a subject's egocentric space and the origin thereof. Subjects usually find themselves at the centre of their world. The tree's in front of us, other people behind us. We may call the way subjects think about the world from their first-person perspective 'thinking about space egocentrically'. In such thinking, things are 'to the left' or 'to the right' of others. Some things are 'in reach' or 'far away'. And here we can also find connections between the thinking subject and the acting subject. In virtue of thinking of the glass of water as egocentrically 'to the left', the subject uses her left arm in order to grab it. Were she to mistakenly think of the glass as egocentrically 'to the right', she would use her right arm and fail to find the glass. This leaves us with the following conception of primitive self-ascription:

PRIMITIVE SELF-ASCRIPTION

When a subject primitively self-ascribes a property, she ascribes that property to the origin of her egocentric space.

The notion of the origin of egocentric space is supposed to provide an explanation of who the subject takes herself to be and thus explain how subjects come to ascribe properties to themselves. We usually know where we are and know what things belong to us and can be moved by our intentions and our will. For instance, my hands are experienced as *mine* because they act according to how I want them to move. And similarly, the feelings of touch or pain in my hands are experienced as *my* pain because these hands form part of my way of directly interacting with the world. But how should we best characterise the concept of the origin of thinking that constitutes us?

My suggestion is that the concept of the origin of egocentric space is in turn illuminated and explained by the notion of the lived body—or

Leib in German. We don't experience ourselves at the geometric origin of egocentric space because we don't experience ourselves as mathematical point-like entities. Instead, we're experiencing the world as being around our lived body with which we can interact with the world. We experience and manipulate the world *through* our lived body.

So, the phenomenological notion of the lived body is central to the account. A subject's origin of egocentric space can be defined via the phenomenological notion of a lived body. On this account, there's a distinction between the lived body and the physical body we call our own. The distinction is probably most plastically illustrated by looking at a partially paralysed subject. While her legs belong to her physical body, they aren't part of her lived body. The subject isn't capable of interacting with the world through her legs because she doesn't feel these legs and can't move them. The lived body is thus constituted through the possibilities of direct interaction with the environment. As I'm typing these words, the movement of my fingers is the realisation of one of the many possibilities of the lived body. The feeling of the keys being pressed down and of the fingers scurrying over the keyboard is an experiential feeling of the lived body.

This leads to the idea that all subjects 'have' a lived body. Strictly speaking, we can't speak of a subject as possessing a lived body. This is because the subject is *constituted* by the lived body while not being identical to it. There couldn't be a subject without a lived body but we can clearly distinguish between the subject and the lived body. It's important to further distinguish between the experienc*ed* physical body and the experienc*ing* lived body. The lived body is the means by which subjects are experiencing and directly interacting with the world around us. And this lived body doesn't necessarily coincide with the physical body that's typically associated with the subject—as we've seen in the case of the paralysed subject. A subject might have an experience of a specific part of her own physical body without thereby taking that part to be part of her lived body. In such a case, the body part is experienced but not experiencing.

Finally, a subject's lived body is essentially first-personal. This is because we can only experience *one* living body—which isn't to say that it's impossible that more than one physical body constitutes the lived body. And this lived body is necessarily experienced as 'one's own'. Edmund Husserl explains this in the following way:

> The perceiving subject perceives 'her' lived body and can, in principle, perceive only *one* lived body as *hers*. She can perceive no other lived body in the peculiar way that she perceives her lived body.[1]
>
> Husserl 1973: 42

If we want to anchor self-ascription in the notion of the origin of egocentric space—and hence the foundation of *de se* thinking in the notion of the lived body—it needs to provide a foolproof connection between the subject and something that she takes to be her own body. The lived body thus understood doesn't require identification of a body because there's only one thing that's phenomenologically given to the subject as the lived body.

This, then, is the origin of our ability to think in the *de se* way. Through primitive self-ascription of properties subjects are capable of *de se* thinking. And the primitive self-ascription of properties consists in the ascription of properties to the lived body. In other words: On the basis of experiencing the world through the lived body—the origin of our egocentric space—we *are* all origins. Thus, the defended account of *de se* thinking can be characterised in the following way: When a subject thinks about herself in the *de se* way she entertains a thought that's partly constituted by her ascribing a property to the lived body. The reason why *de se* thinking is necessarily first-personal is because subjects potentially grasp that they're thinking about themselves. And this kind of grasp is best explained via the phenomenological notion of the lived body.

Now that we have a sketch of the ultimately defended account, we shall first delve into the other possible strategies which attempt to deal with the nature of *de se* thinking and it's peculiarities. This will show us the necessity and the merit of finally opting for the lived body account.

1 This is my translation of the passage: 'Der Wahrnehmende nimmt "seinen" Leib wahr, und prinzipiell kann er nur *einen* Leib als *seinen* wahrnehmen und keinen anderen Leib in der eigentümlichen Weise wie seinen Leib wahrnehmen.'

2

DIVIDE AND CONQUER

> Consider the following case. Dr. Gustav Lauben says, 'I have been wounded'. Leo Peter hears this and remarks some days later, 'Dr. Gustav Lauben has been wounded'. Does this sentence express the same thought as the one Dr. Lauben uttered himself?
>
> Gottlob Frege: 'The Thought: A Logical Inquiry': 297

The logician and philosopher Gottlob Frege was among the forefathers of the Propophile way of thinking. Remember that Propophiles characterise mental states mainly on the basis of a proposition—a possible way the world could be. For instance, a subject's belief that Sydney is the capital of Australia can be individuated by calling on the proposition <Sydney, being the capital of Australia>, which the subject believes to be true. Soon enough the Propophiles realised that the case of *de se* thinking is not as straightforward as they would wish. Frege's quote above expresses a puzzlement that emerges: In one sense, the proposition—what Frege calls a thought—that Alpha holds true in thinking *I've been wounded* is the same as the proposition she holds true in thinking *Alpha has been wounded*. In either case, she pictures the world with Alpha being wounded.

In another sense, Alpha pictures the world in two quite different ways. There's a much more intimate epistemic connection in the case of *de se* thinking which isn't present in the *de re* case. In the former case, she's thinking about *herself* whereas in the latter she's thinking about Alpha, whoever she might be. The fact that these two things happily coincide in the case at hand is a mere stroke of luck—something undetermined by the thoughts themselves. The world happens to be such

that Alpha is also the person who's thinking, but nothing in the two beliefs determines that she's actually thinking about the same thing. Rather, there's room for discrepancy. And thus it seems that the proposition Alpha entertains isn't the same in the two cases. The person who's pictured by believing *I've been wounded* isn't necessarily the same as the person pictured by believing *Alpha has been wounded*.

I've introduced the notion of a proposition as a possible picture of how the world could be. What becomes more and more clear is that this metaphorical way of talking is too vague. This is because we can think of pictures representing reality in quite distinct ways. On the one hand, a picture can represent the world in a very particular and unique way. In this sense, Van Gogh's self portraits are pictures of Van Gogh and no one else. But we can also think of pictures as representing the world in a more general way which could correspond to very different situations. The picture is just a faceless mask which represents anything that fulfils the role in question. In this sense, Picasso's *Three Musicians* could represent many different kinds of musical trios, provided they're composed in a certain way which corresponds to the picture.

Why is this important? Let me explain. Imagine our beliefs would be characterised merely on the basis of their course grained satisfaction conditions. What does it mean for conditions of satisfaction to be course grained? We could say that these conditions answer to the rock bottom question: Which objects need to have which properties in order for the belief to be true? If we then look at the two beliefs from this point of view, Alpha's *de se* belief and her *de re* belief would be identical. They are both true if Alpha—the thing in our world these beliefs refer to—has been wounded. This is the first sense of propositions as very particular and unique ways of representing reality.

But we don't want to say that the two beliefs are identical. After all, Alpha can believe one without believing the other and *vice versa*. So, since Alpha's two beliefs aren't identical, this can't be the whole story. Instead, we need to include a second level: How is the rock bottom state of affairs that Alpha is wounded presented to us in thought? Are we thinking of Alpha on the basis of seeing her directly? Or are we thinking of her on the basis of reading the newspaper headline 'Peaceful protester shot by police'? These are two fundamentally different ways of picturing reality and this insight corresponds to the second way of thinking about propositions. Propositions sometimes leave it

open which particular object is required for the belief to be satisfied. Instead, they care about how things are presented to us in thinking about the world. The truth of the newspaper headline is independent of any particular subject—for instance Alpha—being shot. It's true as long as *some* peaceful protester was shot by the police (A.2.1).

This move from propositions as characterised by their rock bottom satisfaction conditions to propositions as characterised by the way we think about the world in more general terms has some important consequences. It coincides with a specific move of individuating our beliefs and other mental states on the basis of these propositions. There are now two ways of characterising beliefs. On the rock bottom level, we have propositions that correspond to course grained satisfaction conditions. And these are shared by many similar but distinct beliefs. One level above, we're confronted with propositions that are characterised by the different modes of presentation that are present in the way we think about the world and its objects (A.2.2).

Importantly, one and the same object on the rock bottom level can be presented to us once in this way and once in another. For instance, we can think of Wonder Woman either *as* the superheroine wielding the Lasso of Truth or *as* the superheroine wearing indestructible bracelets. In either case, we're thinking about Wonder Woman on the rock bottom level, but this person comes to our minds in different garments. This distinction is important because it's not the same to think that the superheroine wielding the Lasso of Truth is brave and to think that the superheroine wearing indestructible bracelets is brave. After all, these descriptions could potentially be about distinct people in some other possible world. So, our beliefs might correspond to quite different rock bottom conditions of satisfaction. And even if they should coincide on this level, we might not know about this. We can think about Wonder Woman as the Lasso of Truth wielder without knowing anything about her bracelets. If we accurately want to represent the multitude of ways we think about the world, we need to take both this epistemic and the previous semantic difference into account.

Following Frege, this second level of thinking where we bear in mind the different ways of thinking about one and the same thing has often been labelled the *sense* of a belief. Frege observed that a sense alone isn't enough to give us the course grained satisfaction conditions at rock bottom which help determine whether a belief is true or not.

The sense fixes the conditions of satisfaction only in a more open sense which corresponds to how things are given to us in thinking. It doesn't yet tell us which particular objects we're actually thinking about. If we want to know whether it's true that the peaceful protester was shot by the police, we need to know *who* that person was. And that means looking into the world and determining which person is presented to us in thinking *as* the peaceful protestor. Hence, a complete characterisation of our belief needs both this second level of thinking and reference to the specific world we're looking at. Determining how things are presented to us in believing—that's to say, determining the sense of a belief—is only a first, and necessary, step.

It's easiest to illustrate this picture on the basis of an example. Imagine that Alpha is the peaceful protester in our world. Of course, not every peaceful protester is Alpha. Hence, Beta could have been the person we're looking for in a different possible world. If we then take the belief

(6) *The peaceful protester was shot by the police*

and we want to know the course grained satisfaction conditions, we need to know which world we're talking about. The *sense* that's involved in believing (6) doesn't yet determine whether we're looking for Alpha or Beta. It's open to either possibility. But since Alpha is the peaceful protester in the actual world we care about, (6) is true if Alpha was shot. And of course in a different world, the same belief would be true if Beta was shot. Once we've done this detective move, we've got all we need for characterising our mental states. On the one, hand we've got the way things are presented to us in thinking. And on the other hand, we know exactly what needs to be the case in our world for the belief to be true.

Following the Fregean way of laying out the conceptual groundwork, we can distinguish mental states on the basis of these two different levels: course grained satisfaction conditions—or *reference* as Frege called it—and the mode of presentation or sense of the belief in question. The mode of presentation is a functional thing which is associated with the specific mode that an object is presented to a subject *as* the first woman in space or *as* the superheroine wielding the Lasso of Truth or *as* the peaceful protester. And we get from this way of thinking to the intentional object—what our thought is about—of our

thinking by looking into the possible world and finding out who satisfies that description. To take an example, Alpha is the object that 'answers' to the mode of presentation described as 'the peaceful protester' in our possible world.

The following rather technical and abstract picture emerges from this introduction: A belief has a first element which is fully characterised by its course grained satisfaction conditions. Furthermore, a belief has a second element. This second element is characterised by the way the first element or its constituents are presented to us in thinking (A.2.3). We can then say that any two beliefs which differ with regard to one or both of these elements are distinct. So, as in the case of Wonder Woman, two beliefs can have the same first element but different second elements. Believing that the superheroine wielding the Lasso of Truth is brave involves a different second element than believing that the superheroine wearing indestructible bracelets is brave. This is because the way that Wonder Woman is presented to you in either belief is distinct. But, since both beliefs are about the same subject, they have the same first element in our world. They're both *about* Wonder Woman. And hence, the beliefs are true if Wonder Woman is brave in either case.

Both elements are necessary to fully characterise and individuate beliefs. If we only look at the second, element our belief will be underspecified because we don't know the satisfaction conditions of the belief. These conditions are only determined by taking into account the first element of belief. We may ask: In which case is the belief *The superheroine wielding the Lasso of Truth is brave* true? Well, that depends on who that superheroine is and thus which world we're wondering about. After all, the lasso wielder could be a different person in a different world. Hence, we need the first element as well to determine the truth.

But how do we get from the second element of belief to the first? Or, put differently, how do we arrive at the rock bottom satisfaction condition starting from the way things are presented to us in thinking? The answer is quite simple. We ask which possible world we're thinking about and check which objects correspond to our way of thinking in that world. In other words, we take the second element of the belief and 'add' the relevant possible world to get to its first element. For instance, if I want to know what needs to be the case for the belief (6)

to be true, I need to know which possible world we're talking about. If we're talking about our actual world, then Alpha needs to have been shot for the belief to be true. Adding a different possible world might result in a different first element because some other peaceful protester was shot in that world. In such a case, (6) would then be true if that other protester was shot.

This also shows that there's a very direct move from the second element of belief to its first element. Building on Frege's words, we might say that there's a direct and inescapable route from a belief's sense to its reference. Or, more canonically, that sense 'determines' reference. Once we have fixed the second element of thinking and determined which world we want to look at, the dice have been cast. There's only one person that I refer to when I believe (6). And the way that person is presented to me in thinking determines in a given possible world who I'm specifically thinking about. In this sense, the second element 'determines' the first element in any given world.

Of course, two beliefs with the same second element can have distinct first elements but *only* if we're looking at them in two different possible worlds. Once we hold the possible world fixed, we can only think about one particular protester in believing (6). As we saw, the second element isn't specific enough to fully characterise a belief. But as soon as we take the relevant possible world into account there's no wiggle room left for the first element. We then have everything we need to know to determine the belief's truth-value.

We now have a theory that aims to represent the differences in how we think about the world. For that task, it makes use of two elements of thinking. One concerns what the thought is about—giving us the semantic rock bottom conditions of satisfaction, the other is about the way we think about the world. This semantic picture is sometimes called a one-dimensional theory. Unfortunately, this isn't a very intuitive label, so let me explain it. The idea is the following: If we only look at the first element of thinking, there's no room for variation between the belief and what that belief is about—its intentional object. For every belief there's exactly one possible intentional object and *vice versa*. But as soon as we add the second element, there's a dimension along which beliefs about the same intentional object can come apart. We witnessed this when we looked at different ways of thinking about Wonder Woman. One and the same intentional object, but

different beliefs. Hence, the one-dimensional theory adds the dimension of the second element to the fray in order to better represent the differences we find in our mental states.

What the opening quote indicates is that Frege realised that this probably isn't enough to deal with *de se* thinking. The reason is that the uniform relation between the second and the first element collapses in the case of *de se* thinking. I explained that the one-dimensional theory holds that sense determines reference—the second element of thinking determines exactly one first element in every possible world. The relation is between many modes of presentation and one intentional object. But in the case of *de se* thinking it's possible that one and the same type of belief can have many different first elements *in one and the same* possible world. In the one-dimensional theory, that shouldn't be possible. This is because the road from second element to first element is a matter of simply adding a possible world. But in the *de se* case, it isn't enough to hold a possible world fixed in order to know what the first element of the belief *I am tall* is. It isn't sufficient to say: 'Anyone holding that belief in our world is thinking about x' even though that would be sufficient were we to talk about believing (6).

Rather, when we're confronted with *de se* beliefs we need to know more than the possible world we're looking at in order to know who we're thinking about. Most importantly, we need to know who's thinking. If it's Alpha, then the first element concerns Alpha, and if it's Beta, then it concerns Beta. Hence, the one-dimensional move of adding the second element isn't enough to account for *de se* thinking. Instead, we need to divide and conquer and go two-dimensional. We want to be able to make room for another dimension on which we can disentangle the relation between our thinking and the things we think about.

So, let's introduce such a third element next to the two others we've already established into our theory. The job of this new third element is to pin down the *narrow context* of a belief within a possible world. This will allow us to home in on a specific instance of thinking—something which is required in order to understand the peculiarities of thinking in the *de se* way. And once we're equipped with this third element of thinking, we can add it to our arsenal of second and first element to form a complete picture. Our two-dimensional theory can then accurately distinguish between different ways of thinking about an object and all of our thoughts are grounded in some first element of thinking.

This gives us both the diversity required on the level of how the world is presented to us and the rock bottom conditions of satisfaction that semantics requires.

Admittedly, this is all very abstract and sounds convoluted. However, looking at the following temporal belief as a practical example will show the surprising simplicity and usefulness of this way of describing the issue:

(7) *Yesterday it rained.*

If we were still stuck in the one-dimensional theory, we would at some point in our quest to fully characterise this belief ask the question: 'What world are we talking about?' And an answer to that question would supposedly take us from the second element to the first. This would enable us to know what the satisfaction conditions of (7) are. But unfortunately, this move doesn't work in the one-dimensional theory. That's because there's no simple route from the second to the first element in the case of believing (7). Simply inserting a possible world into our one-dimensional functional model won't do. In believing (7), the world is presented to us as a tale of yesterdays. But, we don't know which day we're thinking of when we're thinking about *yesterday*. There are many yesterdays in every possible world, and now we want to know which one we're thinking of. But how?

Obviously, the first element of (7) differs not just between various possible worlds that we can think about. It also depends on the specific context of thinking. Most importantly, it depends on the exact day a subject entertains that belief. If I believed it on Tuesday, the belief would be true if it rained on Monday. And if I believed it on Sunday, the belief would be true if it rained on Saturday. But the second element can't represent this kind of dependency because it only takes into account the possible world we're looking at. It doesn't care about the subtleties of contexts and situation. We introduced it as being constituted by the way we think about the world and we originally thought that only requires taking into account a specific possible world. But beliefs like (7) need a magnifying glass to zoom into the specific time, place and subject of the belief.

That's where the third element of thinking comes into play. Not all of our beliefs are like (6). They don't present us the same thing in different ways in our thinking. Rather, beliefs like (7) are of a kind where

one way of thinking about an object can refer to a myriad of things depending on when, where and who is thinking. The third element of thinking aims to systematically connect these contextual ways of thinking with our by now familiar second element of thinking. Hence, we can say that in the case of (7), the third element is required to pinpoint the varying date that's being picked out by a subject believing it on a certain day.

We thus get the following: The third element of (7) is constituted by a specific day being presented to us as yesterday. Because that day varies depending on when the belief was entertained, the third element doesn't require us to add a possible world. Rather, we need to add the specific context of thinking. In this case, the day of thinking is the relevant part of the context. Hence, the third element of (7) asks for whatever day the belief was entertained and is tantamount to something like 'It rained the day before the subject believed this'. Now, we first need to take a look into the world and determine the context of that particular belief. Importantly, we need to know which day the relevant belief was entertained. So, we identify the context and see that it was Alpha who believed (7) on March, 27, 2016. From that we get the second element 'It rained the day before March, 27, 2016'. Once we have this, we can add the possible world we want to know about and find out what the first element of our belief is. This step is necessary because the description 'the day before March, 27, 2016' could potentially pick out different days in different possible worlds. Maybe March, 26, 2016 isn't followed in all possible worlds by March, 27, 2016. In the actual world, however, the relevant first element is the state of affairs that it rained on March, 26, 2016.

Here's another example of the importance of the third element of thinking which is more relevant to our case of *de se* thinking. Let's determine the different elements of Alpha's belief *I am tall*. Starting from the third element, we get something like 'The subject of this belief is tall'. To get from this to our second element, we have to add the relevant context of thinking. Merely adding a possible world wouldn't yet tell us which belief 'this' belief is. So, let's add the context of the belief and we get a second element which resembles 'The subject believing *I am tall* on June, 1, 2016 in Fribourg is tall'. We still don't know who that subject is, so we need to look into our world and thus determine the first element. Once this is done, we may see that the subject that's

believing herself to be tall on that date and in that location is Alpha. And thus we know the first element of the belief in question. Because Alpha is the subject of the belief we're interested in, we know that Alpha needs to be tall in order for her belief to be true.

We've now characterised beliefs and other mental states on the basis of three elements. The first element is needed to know under which conditions a belief is true. It tells us what a thought is about and gives us the conditions of satisfaction. We determine these conditions by taking into account the second element of thinking. This element concerns the way we think about the first element or its constituents. Since one and the same state of affairs or object can be presented to us in various ways, we need to take a look at this second element in order to do justice to these differences. Finally, we have the third element which is used for beliefs that are dependent on their context (A.2.4). This element is needed because we sometimes think about one and the same thing in various ways and we sometimes think about diverse things in one and the same way. While the second element is adequate for the former task, it's unfit for the latter. This is where the third element shines because it relates one type of thinking with many different second elements.

The introduction of the third element of thinking opens up the following possibility: Beliefs with identical third elements can be about quite distinct first elements even if we hold the possible world we're looking at fixed. Hence, Alpha can think about Monday or Tuesday by believing (7). If she believes it on Tuesday, she's thinking about Monday, and if she believes it on Wednesday, she's thinking about Tuesday. Recall that the same isn't true for the relation between the second element and the first element of thinking. It's a one-way street from the former to the latter. Something similar applies, however, when we move from the third element to the second element. Once we hold a specific situation or context fixed, we'll always get the same second element. The introduction of the third element was therefore necessary because we sometimes think about one intentional object in different ways and sometimes in one specific way about different intentional objects. Our two-dimensional theory allows a move from one kind of belief with its characteristic third element to several different first elements in one and the same possible world—something which wasn't possible in the one-dimensional theory.

Admittedly, this general framework is still very abstract. You might also claim that it involves some overcomplicated moves. In practice, things are much easier because it isn't necessary to consciously go through all these steps in order to determine the intentional object of a belief. However, if we stumble upon an old inscription saying: 'I'm the greatest philosopher of all time' we have to go through these detective moves in order to find out who wrote it down. Usually, we go through the steps seamlessly but sometimes they require more caution. Now, how is this relevant to our goal of understanding *de se* thoughts? In the remainder of the chapter we'll look at three concrete strategies of how we could understand the nature of these different elements of belief. These strategies are meant to be rather general with the mentioned quotes and proponents as illustrations. This somewhat across-the-board kind of approach will naturally gloss over the more intricate details of the specific varieties of theories that are defended under the heading of a particular approach. However, since we want to learn their general advantages and shortcomings, which need to be overcome by an encompassing theory of *de se* thinking, this is a somewhat necessary evil which we condone.

2.1 A FRIENDLY CHARACTER

There's a close resemblance between talking about ourselves by using the word 'I' and thinking about ourselves in the *de se* way. For instance, we can't fail to talk about ourselves when using the first-person pronoun. The reference of the word depends on the context of usage. Additionally, we can find other words that depend on the context such as 'here', 'now', or 'that'. All of these call for a two-dimensional treatment in the semantics of language. This parallel of sensitivity to the context suggests that the way these indexicals and demonstratives work in natural languages like English or Breton might serve as a key and blueprint to understand *de se* thoughts. The idea is to start from our understanding of how we—as language speaking subjects—talk about ourselves. If we understand how we come to use words such as the first-person pronoun in a meaningful way, we might learn something about our capability to think about ourselves in the *de se* way. I'll call this approach the *linguistic* approach. For our purposes, we can characterise the linguistic approach as claiming that we can understand how

language using subjects think about themselves *via* their use of the first-person pronoun in speech.

In his 'Demonstratives' (1989), Kaplan provides us with a very characteristic version of this kind of approach. We can use his account as a kind of template to examine the linguistic path to *de se* thinking. Kaplan wants to find out 'what is said' by some peculiar kinds of utterances like 'That's green' or 'I'm tall'. In this context, he draws a distinction between the *character* and the *content* of an expression because he notices that a purely one-dimensional semantic theory won't suffice for these utterances. Demonstrative expressions like 'that' are highly dependent on the context of their utterance and need a two-dimensional treatment instead. We can point out many different things using the word 'that' and only the context will determine which one is pointed out. Merely adding a possible world to the picture tells us practically nothing about what particular thing we're picking out with that expression. To get the referent of such a demonstrative we need to know the context of the utterance.

The general layout of Kaplan's two-dimensional semantic theory of linguistic expressions is as follows: The role of what I called the third element of thinking is played by the Kaplanian *character* of an expression or utterance. And what I called the second element is akin to the Kaplanian *content*. At rock bottom—our first element—we then have the appropriate *extension* for the relevant expression. For instance, if we're looking at a sentence, the rock bottom extension is a truth-value, and if we're looking at a term like 'the best female climber', it's an individual.

Let's look at these three elements within the linguistic approach in a bit more detail. The content of an expression is what's being evaluated for its truth in a given circumstance. We want to have something that can be either true or false simply by looking whether it's true in a given world. It should be possible to evaluate the truth of a content in any given possible world. For instance, the content of my statement 'The first woman in space just turned 80' is that the first woman in space just turned 80. This can be evaluated in our world for its truth by determining who the first woman in space was and checking whether she just turned 80 or not. Here, we don't need to be bothered by the more fine-grained context of the utterance. Whoever will utter that sentence in our world will thereby talk about Valentina Tereshkova. Hence, we

get from the content of an expression to its truth-value by evaluating the content in a specific possible world. The resemblance between the content and the second element of thinking becomes very striking. Furthermore, the relation between content and extension is similar to the relation between the second and first element: The content of an utterance 'determines' the extension in every possible world. In other words: The relation between the content and the extension is many to one. One extension can be picked out by many contents.

But it isn't always this easy to establish the content of an utterance. Sometimes we need some additional information to get to something which can be evaluated for its truth in a world. In Kaplan's theory, that job is reserved for the character of an expression. It takes us from a specific context of utterance to a content which can then be evaluated for its truth in a possible world. In the case of saying 'It rained yesterday', the expression 'yesterday' is such that we have to take into account the context of utterance in order to know its content. Here's where the character comes into play. The character of 'yesterday' is such that it picks out the day before the utterance in order to form the right content. Once we have that, we can evaluate the content in the world we're interested in—usually our own actual one. Expressions such as 'yesterday' require a two-dimensional treatment—and thus a character—because we don't automatically know its content just by looking at it. We don't immediately know which day we're talking about. It heavily depends on the context, and not just the possible world we're looking at, which day is picked out by a specific use of the expression.

All expressions have a specific character. But only some of them have a character that factors in the context of the utterance within a specific possible world. Because our conversations are usually anchored in one and the same world, the expression 'the first woman in space' has a fixed character in normal circumstances. It refers to Valentina Tereshkova in all possible contexts in the actual world. Hence, we don't need to take a closer look at the character and context in order to attain an evaluable content. On the other hand, the expression 'I' has a character that takes into account who's using it. It doesn't always refer to the same individual in all possible contexts. We can use one and the same expression to talk about many different subjects. So, we need to bear in mind the context of utterance in order to know what's being said by the speaking subject.

Of course, what interests us the most is how the linguistic approach characterises the nature of the third element—the character. After all, this is what forms the crucial step for dealing with *de se* attitudes and decides the fate of the theory for our purposes. Characters aren't just functional devices that take us from a context to a content. They have to 'do' this somehow. What is it about the character of an expression that does this job? As explained, characters are fundamentally explicated in functional terms in pretty much the same way as I characterised the third element of thinking in general. While contents are functions from possible worlds to extensions, characters are functions from possible contexts to contents. Or, to put it differently, we get from a specific content to an extension by 'adding' a possible world, and we get from a character to a content by 'adding' a context. So, if we want to know what a subject says by uttering 'I'm tall', we look at the characters of the contained expressions and the context of utterance to arrive at a content. Once this is done, we can evaluate that content for its truth in a specific possible world. But what determines the function of a character? For Kaplan, it's the rules of language:

> The character of an expression is set by linguistic conventions and, in turn, determines the content of the expression in every context. Because character is what is set by linguistic conventions, it is natural to think of it as *meaning* in the sense of what is known by the competent language user.
>
> Kaplan 1989: 505

Speakers of English know that we always refer to ourselves when we use the word 'I' and related expressions. After all, that's the meaning of the first-person pronoun; it's how we use that expression in speech. Our linguistic conventions thus determine the character of the innocent word 'I': In all contexts of utterance it refers to the subject using that expression. So, when we're confronted with a specific instance of that expression we can determine the content on the basis of its character and the context. Imagine I sit across from you and you tell me 'I'm thirsty'. On the basis of my knowledge of the character of 'I', I could establish that the speaker of that particular sentence is thirsty. I then identify the context of your utterance and determine that the content is something like the proposition <the subject uttering 'I'm thirsty' on

July, 7, 2016 in Fribourg, being thirsty>. Now, I have a proposition that I can evaluate for its truth in our world.

A different example involving an expression with a variable character would be Alpha uttering 'That tennis player is sublime'. Speakers of English know that the reference of 'that' is determined by some kind of demonstration—pointing out who the subject intends to talk about and refer to. Of course, such a demonstration is heavily dependent on the context. After all, we can point out many different things with just one finger. Sitting in the ranks of Centre Court in Wimbledon, you see that Alpha points out the player to her left. You can thus establish that the content of her utterance is something like <the tennis player to the left of Alpha on Centre Court in Wimbledon, being sublime>. That again can be evaluated for its truth. If that player is Steffi Graf, then Alpha's utterance is true if Steffi Graf is sublime. Again, this sounds more convoluted than we're used to in everyday talk. However, the cases where speakers aren't easily identified or demonstrations are unclear show that this complex two-dimensional semantic system is necessary and always works in the background. It's just very well internalised in our everyday speech.

Here's an example of an expression with a character that doesn't seem to require us to take the context of utterance into account. The meaning of 'the element with the atomic number 79' is usually such that it picks out one and the same element in all contexts within a possible world. It isn't the case that the expression refers to gold on Monday and helium on Wednesday. It also isn't the case that it refers to different elements when used in Greenland as opposed to Samoa. And it also doesn't matter whether I'm using the expression or Valentina Tereshkova is. We're always talking about gold. Of course, in a different possible world, some other element might have the atomic number 79 and thus be the extension of that expression. But that variance is already covered by the content which allows us to pick out different things in different possible worlds. And because the content of the expression doesn't vary across contexts, the character is fixed.

So, Kaplan identifies the character of an expression with its conventional meaning. But what exactly does that entail? Notice first that conventional meaning is a normative and not merely a descriptive term. The meaning of a word tells us how we *should* use that word in a given circumstance and not just how we actually use it. If we want others

to understand what we're saying, we have to follow some rules of language. For otherwise, we might not be able to communicate what we're thinking to others. Of course, nothing stops you from using the expression 'I' to refer to Mont Blanc. But you'll have a hard time getting your point across if you're not using an expression according to its conventional meaning.

Another important feature is that the conventional meaning of an expression is quite rigid. It might change slightly over time but it usually stays more or less the same. To illustrate this, let's look at the meaning of the first-person pronoun: It always refers to the speaker. So, when Alpha utters 'I'm tall', she's using the first-person pronoun with the same meaning as when Beta says 'I'm tall'. Both statements use the same words and have the same meaning. Even if Beta meant to convey that she's tall in Breton, her words would have a similar meaning as Alpha's English utterance. The fact that the character is established by the conventional meaning explains why Alpha and Beta were, in an important respect, saying the same thing despite talking about two completely different subjects. They both use the expression 'I' which has the same character, and thus the same meaning, in both cases. Due to the two-dimensional nature of that expression, it allows us to refer to quite different things despite having the same conventional meaning in each case.

How is the meaning of an expression determined? Kaplan holds that competent language using subjects have some *knowledge* of the rules of language—after all, that's what makes them competent. They know that names refer to the things they name, they know that predicates can be applied to various things, they know that some expressions can be used in various contexts to refer to different things. This kind of knowledge establishes the character of the first-person pronoun. We, as competent language users, know that we use that expression to refer to ourselves in all cases. In other words, 'I' always refers to the speaker. Our knowledge of how language is to be used thus determines the character of the expressions we use.

Now, there's a final element in Kaplan's linguistic approach which is of great importance when we think about the role of *de se* thinking. That the meaning of an expression like 'I' is determined by what is known by a competent language user doesn't imply that a subject needs to be *aware* of these rules if she utters 'I' with meaning. After

all, we can utter meaningful words without understanding them properly. But in the *de se* case, we want more than that. We want subjects to talk about themselves *knowingly*. For this task we need to add some epistemic element which connects our use of language with our capacity to think in the *de se* way. Subjects need to be aware of the rules of language which govern the use of the first-person pronoun. For otherwise, they wouldn't be capable of using the expression 'I' to express their *de se* thoughts. Instead, they would mindlessly utter meaningful words. But because of their lack of knowledge, these words and utterances wouldn't serve to express what they usually do. The point is the following: For an expression to play the role in language that it's supposed to play, users need to employ that expression somewhat knowingly. So, a subject needs to have some kind of knowledge that her use of the first-person pronoun refers to *herself* in her own use if she wants to successfully use that expression.

Imagine a parrot who supposedly has no idea of how language works. Such a subject can't express her thought that she's tall by uttering 'I'm tall' even if that would be the right way to express her thought were she aware of the conventions of the English language. That's because she has no idea that uttering 'I'm tall' is the right way to express your belief that you're tall. And this ignorance is also responsible for the fact that if an incompetent speaker like our parrot were to utter such a sentence, that utterance wouldn't express what is normally expresses—even if *we* can determine the conventional meaning of the utterance.

Kaplan addresses this important aspect of the character only quite late in his paper when he discusses some epistemological ramifications of the theory. This is quite surprising. After all, our knowledge of the rules of language is necessary if we want to use it to express what we're thinking. We're not babbling like infants after all. And since the linguistic approach to *de se* thinking wants to explain our capacity to think about ourselves via our ability to talk about ourselves, it should absolutely include such an epistemic element which ensures the knowing use. In the end, though, Kaplan delivers:

> What we must do is disentangle two epistemological notions: *the objects of thought* and the *cognitive significance of an object of thought*. As has been noted above, a character may be likened to a manner of presentation of a content.

> This suggests that we identify objects of thought with contents and the cognitive significance of such objects with characters.
>
> Kaplan 1989: 530

What's the argument for the claim that the cognitive significance lies in the character of an expression? Here's one possible answer. It seems that we can meaningfully wonder 'Is this me?' while looking into the mirror. And for that to be possible it certainly isn't necessary to know the Kaplanian content or extension of all the terms involved. After all, part of what we're wondering about is whether the extension of 'this' and the extension of 'me' are identical. Thus, it seems that we need to pin down the epistemic aspect of thinking in the character and not the content of an expression. The cognitive significance is rooted in the character of an expression because two fundamentally different objects can be given to us in very much the same way. And in two situations where two distinct objects are given to us in the same way, we act identically. For instance, most subjects will try to get out of danger when they utter to themselves 'I'm in danger'. That's because they're all being presented to themselves on the basis of their use of the first-person pronoun with its character.

This quote reflects what's special about *de se* thinking: the way we're presented to ourselves in thought. It's a way that's fundamentally different from the way we think about other things in the world. And with it comes a certain cognitive significance that's distinct from the significance of thinking about ourselves in other ways. Alpha will draw different conclusions from her *de se* belief *I'm tall* than from her *de re* belief *Alpha is tall*. These two respective beliefs play fundamentally different roles in her reasoning and motivation. And this difference is reflected in the different expressions we use to report our beliefs. The expression 'I' has a different character than the expression 'Alpha'. This isn't merely because they are different kinds of expressions with different semantical properties. It's also because they are used to express beliefs with different epistemic roles in reasoning.

Now, let's recapitulate the linguistic approach. Remember that it claims that we can understand how language using subjects think about themselves in the *de se* way *via* their use of the first-person pronoun in speech:

LINGUISTIC APPROACH TO DE SE THINKING
When a language using subject thinks about herself in the *de se* way, she entertains a thought that's constituted by her knowledge of the character of the first-person pronoun which is governed by the meaning rule that 'I' always refers to the speaker.

The linguistic approach thus converges on the characterisation of *de se* thinking via the linguistic meaning of the first-person pronoun and a subject's *grasp* of that meaning. The meaning is then identified with the character of an expression according to the two-dimensional semantic theory. Unsurprisingly, this move nicely accounts for the semantic features of *de se* thinking. The interaction of character, content, and extension makes sure that our *de se* thoughts are always about the thinking subject and that the satisfaction conditions systematically depend on the subject that's entertaining a *de se* thought.

However, the approach also needs and wants to account for the epistemic features we've identified. This is why an epistemic constraint was introduced to the theory. The subject can't merely comply to the rules of meaning which govern the use of the first-person pronoun. If we want her speech to be a proper expression of her thought, she also has to do so knowingly, for otherwise her use wouldn't be endowed with the cognitive significance that's characteristic of *de se* thinking.

In other words: If a subject doesn't know that her use of 'I' always refers to the speaker, she wouldn't be in a state of thinking endowed with the characteristic epistemic features of *de se* thinking. As a consequence, this provides us with a necessary minimal requirement on the contribution of the character of the first-person pronoun to *de se* thinking. If language is supposed to be explanatory of our abilities to think in the *de se* way, then a subject has to be aware of the fact that she refers to *herself* by using the first-person pronoun. Without this epistemic contribution, the linguistic approach doesn't get off the ground as a worthwhile candidate. After all, the epistemic features of *de se* thinking form the heart of the explanatory project we're pursuing.

You might now object that such an approach is quite limited since it only attempts to explain how language using subjects think about themselves in the *de se* way. What about non-linguistic subjects like koalas and cheetahs? Ultimately, we couldn't use such an approach to explain how non-linguistic animals might think *de se* thoughts be-

cause they don't express their mental states linguistically. If you have this objection, I'm conceding it to you. The approach is indeed very narrow and unsuited to give an account of this kind of non-linguistic *de se* thinking. Even if there are arguments against the presence of *de se* thinking in nonhumans, this claim should be supported independently of the linguistic approach. Nonetheless, there's a reason to take the linguistic approach seriously. That's because a proponent of the approach might counter in the following way: It's unclear whether we can even find *de se* thinking in non-linguistic subjects. However, it's very clear that we find it in language users. And if we look at the phenomenon from the perspective of language use, we at least learn something about how language users, such as human subjects, think about themselves in the *de se* way. Once this is achieved, we have a better understanding of the nature of *de se* thinking and might still broaden our theoretical scope to include non-linguistic subjects.

We now have a dispute between those who want to achieve a neutral account of the nature of *de se* which leaves room for non-linguistic thinking about oneself on the one side and those who hold that our only reliable tool to uncover what a subject is thinking is the subject's linguistic behaviour on the other. My reaction is to sympathise with the first group but refrain from providing a justified decision between the two for the moment. Maybe there's no need to pick a side because the linguistic approach doesn't even account for *de se* thinking in language users. Indeed, that's exactly what we'll see. I want to discuss two crucial problems for the approach. And these remain valid even if we narrow the approach down to how language using subjects think about themselves in the *de se* way.

The first concerns a certain paradox of self-consciousness. How can a subject learn and understand the meaning rule of the first-person pronoun if she isn't already capable of some non-linguistic form of *de se* thinking? Isn't the meaning rule dependent on a prior knowledge of how we think about ourselves? Hence, isn't the direction of explanation exactly the opposite of what the linguistic approach holds? The second problem asks how the approach explains the nature of intentional action. Is the character of the first-person pronoun capable of doing justice to the motivational force of *de se* thinking? Does knowledge of the character of 'I' play the right kind of practical role in order to motivate a subject to act according to her own reasons?

Let's first look at the problem concerning the way we learn and apply the meaning rule for 'I'. On the linguistic approach, our knowledge of the character of the first-person pronoun is supposed to account for our capability to think about ourselves in the *de se* way. It's because we're competent users of the expression 'I' that we're able to entertain beliefs like *I'm tall* or *I'm hungry*. But we can now ask: How can a subject acquire knowledge of the meaning rule for 'I'? The supposed answer is that a subject already has to have the capacity to think in the *de se* way in order to come to understand the meaning rule. If I want to linguistically express that I'm thinking about myself in the *de se* way, I need to have that capacity in the first place. That's because I'm only capable of grasping how to use the first-person pronoun according to its meaning rule if I can think about myself in the *de se* way already. José Luis Bermúdez discusses this problem in his *The Paradox of Self-Consciousness* (1998). He explains the dependence between knowing the character of 'I' and *de se* thinking in the following way:

> The point here is that the capacity for reflexive self-reference by means of the first-person pronoun presupposes the capacity to think thoughts with first-person contents, and hence cannot be deployed to explain that capacity. In other words, a degree of self-consciousness is required to master the use of the first-person pronoun.
>
> Bermúdez 1998: 18

The linguistic approach holds that we explain our capacity to think in the *de se* way by our knowledge of the character of 'I'. But Bermúdez wants to argue that this knowledge is only possible if we're capable of thinking *de se* thoughts already. Without the knowledge that uttering 'I' refers to *yourself*—a piece of *de se* thinking—you haven't fully grasped the meaning of the first-person pronoun. He concludes that we can't use our linguistic ability to explain *de se* thinking. If we take the explanatory route of the linguistic approach, we're going in circles. The capacity we want to explain is itself required for the capacity we think does the explaining.

How does Bermúdez argue that *de se* thinking is required for the mastery of the first-person pronoun? As we saw in the elaboration of Kaplan's account, some knowledge on the part of the subject is required to do justice to the epistemic aspect of *de se* thinking. A subject

can use the first-person pronoun in order to express her *de se* beliefs only if she knows that by using 'I' she's thereby referring to herself. But this kind of knowledge presupposes her ability to think in the *de se* way. For otherwise, she wouldn't understand that she refers to *herself* when she's using the first-person pronoun. She might only grasp that she's referring to the speaker—whoever she might be. But grasping that you *yourself* are the speaker—and thus the appropriate referent of your use of 'I'—requires you to think about yourself in the *de se* way. Or, as Bermúdez puts it: 'Employing a token of the first-person pronoun in a way that reflects mastery of its semantics requires knowing that one is the producer of the relevant token and that is a piece of knowledge with a first-person content' (Bermúdez 1998: 15–16). Since first-person contents are *de se* in nature, our mastery of the first-person pronoun requires the capacity to think in the *de se* way.

So, Bermúdez' argument is indeed quite simple. A subject's grasp of the meaning rule of the first-person pronoun requires her knowledge that her use of 'I' refers to *herself*. This kind of knowledge is *de se* in nature. Therefore, the acquisition and application of the meaning rule for 'I' depends on a prior capacity for *de se* thinking. And this means that learning and employing the meaning rule of the first-person pronoun isn't the basic explanatory capacity we're after. Hence, the linguistic approach to *de se* thinking fails (A.2.5).

Now, it's important to note that this argument doesn't imply that the Kaplanian character of 'I' is ill-defined. There's indeed nothing wrong with the meaning rule for the first-person pronoun. Rather, it's the proposed explanatory project of the linguistic approach which gets us into circularity troubles. The approach is circular only because it holds that the ability to think in the *de se* way has to be explained on the basis of a subject's knowledge of the character of 'I'. More specifically, the trouble arises with the addition of the epistemic constraint. Remember, this constraint was introduced to exclude accidental self-reference. Thinking about oneself in the *de se* way is a way of genuinely and necessarily thinking about oneself. So, the use of the first-person pronoun has to make sure that the subject genuinely thinks about herself in the *de se* way. This has to exclude some accidental way of thinking about oneself. If the speaker grasps that 'I' always refers to the speaker, but fails to understand that she herself is the speaker, she only referred to herself by accident through her use of 'I'. She didn't intend to speak

about herself. On the linguistic approach, the subject excludes this accidental self-reference only through her *knowledge* of the character of 'I'. However, in order for this knowledge to be of the kind that secures the kind of reflexive self-reference we want, the subject needs to anchor it in her already present ability to think *de se* thoughts.

The downfall of the linguistic approach thus arises from its move beyond the mere semantic realm. As a semantic account of the reference of 'I', it doesn't face any vicious circularity. It's the claim to serve as an explanation of the possibility of *de se* thinking—with all its epistemic features—which leads to futility. As soon as we build epistemic requirements on top of the semantic account, we topple the explanatory project because we take for granted what we set out to explain.

That's already really bad news for the linguistic approach. But two is better than one, so let's reinforce our dismissal of the linguistic approach with a second problem. I argued that there's a special relation between *de se* thinking and the ability for intentional action. I argued that subjects need to take reasons for certain actions as their *own* reasons in order to be motivated to act on them. And this in turn requires them to think about their own reasons in the *de se* way as *their own*.

So, when Alpha is being attacked by a bear, there's a strong reason for her to run away. She takes this reason as her own by believing *I'm being attacked by a bear* and maybe concludes from that that she's in danger. Through this, she takes the reason that Alpha is in danger as a reason for *herself*. And she does this because that reason is presented to her in the *de se* way. Were it presented to her in some other way, she wouldn't necessarily be motivated to act on that reason without some additional *de se* belief.

Hence, our favoured account of *de se* thinking needs to be able to do justice to this feature. Can the linguistic approach succeed? Kaplan's proposal for the linguistic character of 'I' is roughly tantamount to 'the speaker of this sentence'. If we want to understand what a subject means by uttering 'I'm in danger', we should substitute the character of 'I' accordingly and get:

(8) 'The speaker of this sentence is in danger.'

However, this certainly doesn't express what the subject thinks when she's entertaining the *de se* belief that she herself is in danger. Why not? Because there's an epistemic problem. It's perfectly possible for Alpha

to utter (8) and wonder whether she herself is the speaker of (8). In other words, she can—from her point of view—reasonably utter 'The speaker of this sentence is in danger *but* I'm doing fine'. Understanding the character of 'I' results in the insight that the expression means 'the speaker of this sentence'. But this results in an unfortunate elimination of the crucial *de se* element. We lose the grasp that Alpha is talking about herself.

The Kaplanian semantic theory eliminates the *de se* element from the picture by mischance. But that's the opposite of what we want. We don't want to eliminate the *de se* because that would make the desired features hard or impossible to explain. Indeed, we've already witnessed in our prior objection the unfortunate interaction between two elements of the linguistic approach. On the one hand, we characterise *de se* thinking through the knowledge of the character of 'I'. On the other hand, we eliminate the *de se* element through our two-dimensional semantic theory of characters. This makes it hard to explain how a subject takes her belief that she's in danger to be a reason for her to run away.

If the character of 'I' is to always refer to the speaker, the *de se* belief *I'm in danger* seems to amount to the acceptance of the sentence (8). But why should a subject react in the usual way to the danger she's in when she accepts (8)? Why should she, upon her acceptance, run away? If a subject accepts that the speaker of a certain sentence is in danger, that doesn't usually lead to her running away. Something crucial is missing: She has to believe that she *herself* is the speaker. But that just reintroduces a new *de se* belief which requires two-dimensional analysis. And if we analyse this new belief in the same way as the original one, we inevitably end up in an infinite regress. Something that's better avoided.

We can now see that the linguistic approach is bound to failure because it pins down the issue of *de se* thinking on a level that's not basic enough for our purposes. From within the Kaplanian picture, we need a rather demanding notion of character if we want to do justice to the features of our capacity to think about ourselves in the *de se* way. This notion of character requires extensive knowledge by the subject in order to do the required epistemic job. Unfortunately, this will inevitably result in a notion that demands too much epistemic work from the subject to get off the ground. What we want to explain—the capacity for

de se thinking—only really works if the subject is already capable of *de se* thinking. So, we might be better off with something more basic to account for the specific work that *de se* thinking provides.

2.2 DIGGING FOR THE FUNDAMENT

As we've seen already, there are certain relations between *de se* thinking on the one hand and self-knowledge, reasoning, and rationality on the other. But what's responsible for these connections? We might try to forge the necessary links via our linguistic competences. We could argue that subjects are rational because they assent to the appropriate sentences on the basis of accepting other sentences and understanding how they logically relate to others. And by using the first-person pronoun in the appropriate circumstances, they express their self-knowledge in a way that makes others' doubt void.

However, I argued that this approach leads to some devastating problems in the case of *de se* thinking. Circularity and infinite regress loom around every corner. Not only that, it seems that we have to look at something more basic than language in order to pin down what's special about *de se* thinking. Our ability to think about ourselves is more ancient and basic than our use of language. And our competence of the first-person pronoun seems to be grounded in a prior capacity to think in the *de se* way. It's only because subjects have some other capacity that they're able to rationally assent to sentences and express their self-knowledge linguistically. But what might that capacity be? Maybe we shouldn't look at our mastery of the first-person pronoun but rather at the kinds of *concept* that are employed in thought. More precisely, the *first person* concept seems to be a good place to start if we want to better understand *de se* thinking.

Let's call this the *conceptual* approach to *de se* thinking. In its most general formulation it holds that we think about ourselves in the *de se* way through employing the first person concept in thinking. So, whenever a subject entertains a *de se* thought, she's thereby using that concept in her belief, judgement, or desire. The first person concept thus plays the semantic, epistemic, and pragmatic roles that are required of our way of thinking about ourselves. It's only because we conceptualise ourselves using the first person concept that our *de se* thoughts have their special qualities.

In order to better understand and assess this approach, we should first ask what a concept is in general. As usual, there are a multitude of theories that elucidate the nature of concepts. Some philosophers think that concepts are merely the discriminative elements of thinking that categorise the unstructured perceptual input for our minds to process. Others have a more elaborate and demanding theory of concepts which draws connections to rationality and reference. Still others might hold that learning a concept is nothing more than learning how to use the relevant word in one's language.

In the context of our two-dimensional journey we'll take a closer look at a theory which builds on the self-referential nature of the first person concept. This way of treating concepts is heavily influenced by Frege. From him we get the idea that the first person concept is such that it refers to the thinking subject in all possible situations and contexts. Other concepts aren't like that. Their referential rules don't mention the circumstances of thinking in the same way. For instance, the reference of the concept of an astronaut doesn't shift when different subjects employ it in thought. So, the first person concept is different from other concepts in that its reference depends heavily on the context of thinking. And this is also the reason why it requires a two-dimensional treatment.

Many concepts do justice to the fact that we can think about one and the same things in different ways—reminiscent of the second element of thinking. We can think of Jane Austen as *the author of Pride and Prejudice* or as *the most famous female novelist of the 18th century*. But since we can also think in one and the same way about distinct things—such as when we're thinking about yesterday—our theory of concepts has to account for that as well. This is why concepts sometimes involve an element that takes the context into account.

The conceptual theory we're focusing on builds on a Fregean understanding of concepts which individuates them on the basis of two criteria: their contribution to the conditions of satisfaction and the way they present objects in thinking. On such a view, concepts are partly constituted by a specific way of thinking about something and form part of complete judgements. If they are applied in thought, they're partly applied on the basis of one's grasp of the rules of reference of the concept in question. So, when Alpha believes (6), she applies the concept of a protester. This concept is partly constituted by the rules

of reference. They tell us when something falls under the concept and when it doesn't. On the other hand, it's constituted by the way we think about these objects—as protesters and not as the ordinary subjects that they also are. Of course, these two things are often interrelated and don't need to be independent at all. Alpha's application of the concept of a protester in her belief (6) is thus partly due to her grasp of the rules of reference of that concept and partly because she thinks of someone *as* a protester (A.2.6).

The neo-Fregean philosopher Christopher Peacocke provides a nice example of this approach to concepts in general and the first person concept in particular. In his *The Mirror of the World* (2014) he explains:

> I will be taking it that a concept is individuated by its fundamental rule of reference. Intuitively, the fundamental rule of reference for a concept states the condition that makes something the reference of the concept. … The psychological significance of the fundamental rule of reference for a concept lies in part at least in the contribution it makes, in combination with the rest of a thinker's information, in helping to determine what are, and what are not, good reasons for judging contents containing the concept.
>
> Peacocke 2014: 81–82

We can nicely observe the two individuating criteria we've illustrated above in this passage. On the one hand, the fundamental rule of reference is a semantic rule that determines which things fall under the concept. They tell us what needs to be the case for something to be referred to on the basis of a use of the concept. On the other hand, a subject who applies the concept in thought has to grasp this fundamental rule of reference so that she only employs the concept in the appropriate judgements.

Now, how does this play out in a concrete example? On the basis of this exposition, we might say that Alpha and Beta apply the same concept of danger when they individually believe:

(9) *The spider is dangerous.*

This is because they would draw the same inferences—provided they're otherwise sufficiently similar—on the basis of believing (9). And they draw the same inferences because they both grasp that the

concept of danger applies to active volcanoes, poisonous substances, and corrupt politicians, but not to napkins, empty water bottles, and toddlers. In other words, they both associate the same fundamental rule of reference with the concept of danger in believing (9). If they separated all things in the world into two piles, one labeled 'dangerous' and the other 'not dangerous', their piles would be sufficiently similar. These fundamental rules also incorporate the reasons a subject has to judge that something is dangerous.

How do we as outsiders know if two subjects apply the same concept in believing (9)? Well, imagine Alpha reasons that the spider is dangerous, therefore she should be careful. Meanwhile, Beta ponders that the spider is dangerous, therefore it's coloured. Other things being equal, the concept of danger—the one that applies to active volcanoes, poisonous substances, and corrupt politicians—doesn't have a lot to do with colour, so it shouldn't be brought into rational connection with that other concept. The reasons for judging something to be coloured *prima facie* have nothing to do with the reasons one has to judge something as dangerous and *vice versa*. But, something being dangerous is a good reason to judge that one should be careful.

Hence, whatever concept Beta applied in believing (9), she wasn't applying the regular one of danger. And that's because she draws fundamentally different inferences on the basis of thinking of the spider *as* dangerous. This again points to the fact that she associates a fundamentally different rule of reference with her concept of danger. And since concepts are individuated by their rules of reference, Alpha and Beta aren't applying the same concept in their beliefs. Were they to separate all things in the world into two piles, they wouldn't be similar at all. And were they to think about when we have good reason to think of something as dangerous, they would significantly disagree.

Now that we know a little more about how concepts work in general, we can apply the story to the first person concept and see whether the conceptual approach is satisfactory for our purposes. Let's start with the fundamental rule of reference that individuates the first person concept. After all, the reference rule plays a crucial role in specifying the nature of a concept and in characterising how something is given to subjects in thought.

Luckily, Peacocke offers us an account of the first person concept. He holds that the fundamental rule of reference for the first person

concept is the following: Whenever the first person concept is applied in an instance of thinking by a specific subject, it refers to that specific thinking subject in virtue of her being the producer of that event of thinking. Or, to formulate it in a more technical and convoluted way: Whenever there's an event of thinking and a thinking subject applies the first person concept in that specific event of thinking, then the concept which the subject used in her event of thinking refers to that specific thinking subject in virtue of her being the thinking subject of that particular event of thinking.

Here's an example of that account in action. When Alpha believes *I'm tall*, she thereby applies the first person concept in an event of thinking. The event of thinking is her forming the belief that she's tall. And since she's the producer of that event, she's also the referent of that application of the first person concept in virtue of the fundamental rule of reference. So, she's thinking about herself in applying the first person concept because every application of the first person concept refers to the thinking subject. The same applies to Beta when she believes *I'm tall*. The fundamental rules of reference of the first person concept represent the reflexivity we've already witnessed.

Is this semantic analysis enough to account for our features of *de se* thinking? Well, the semantic part of the linguistic approach didn't do the job before. But the semantics of the conceptual approach are quite similar to what we had in the linguistic case. We saw that semantics is one thing, but *de se* thinking requires some epistemic flavour too. In the linguistic approach, we had to include a subject's knowledge of the fact that 'I' always refers to the speaker. And something similar is required for the conceptual approach.

The subject needs to have some epistemic access to the fundamental rule of reference that individuates a concept. Peacocke is well aware of the need for such a qualification when he writes that 'a theory of concepts needs to be accompanied by a theory of what it is to grasp those concepts'. And he provides such a theory by explaining that 'grasp of a concept consists in tacit knowledge of its fundamental reference rule' (Peacocke 2014: 84). What does that amount to? When a subject makes a judgement, she needs to employ the required concepts. And that requires some *grasp* of the concepts in question—she has to somehow be *aware* of what she's judging and the concepts she uses therein. Grasping a concept needn't be a conscious activity but rather consists

in some form of knowledge of what it means to apply the concept in thinking. She needs to be aware of what are and what aren't good reasons to judge something to fall under a concept.

Exactly what a subject's *tacit* knowledge of the fundamental rules of reference generally amounts to isn't quite clear though. And even Peacocke, unfortunately, leaves us somewhat in the dark. There are at least two apparent options which attempt to explain how a subject possesses tacit knowledge of the rules of reference of a concept. One possibility is that the tacit knowledge simply manifests itself in the subject's correct application of the concept in her beliefs and inferences. Another alternative is that it's based on some other underlying epistemic capacity which forms the basis of the concept.

Let's pause for a moment and assess these two options. When we look at the first option, we're left wondering what explains the subject's correct application. What's responsible for her applying the concept in the right way? It can't magically happen to be that way. For example, we may ask why Alpha apparently applies the concept in the right way in believing (9) and why Beta doesn't do so. Something in their application must indicate that only Alpha applies the concept *correctly*. But, we can't offer an account that uses tacit knowledge of the reference rules of a *different* concept because that just shifts the problem over. Now we would need to explain what it is to grasp that other concept. What we need is some independent ground on which we can decide whether a subject grasps a given concept or not.

This leads us to the second option. Maybe there's some underlying capacity which establishes and enables the grasp of a concept. The tacit knowledge would then be established in a more basic capacity. One possibility could be connected to the kinds of reasons that subjects take to be good reasons for judging something to fall under a concept. But if that underlying capacity isn't the grasp of some other concept—but rather something non-conceptual—it seems to jeopardise the conceptual approach as a whole. After all, that approach aims to explain the characteristic features of *de se* thinking on the basis of a subject's use of the first person concept. We don't want some non-conceptual ability of the subject to meddle with this endeavour. If the important work in explaining what's special about *de se* thinking isn't done by the first person concept but rather some more basic capacity, then we can reasonably ask what explanatory work the conceptual approach provides.

The conceptual approach tells us that the first person concept is responsible for the special features of *de se* thinking. But our analysis of the notion of tacit knowledge places doubt on this claim because it shows the requirement of something more basic. After all, we want to know what gives a subject a reason to apply the first person concept in *de se* thinking.

There's a real difficulty of explaining what exactly tacit knowledge amounts to. And we can't merely brush the question and the requirement of tacit knowledge aside. This is because the knowledge constraint is necessary for two reasons. First, the mere semantics of concepts doesn't account for all our features of *de se* thinking. We need some epistemic element in order to account for things like self-knowledge and immunity. Secondly, we want the knowledge to be *tacit* to avoid the implausible assumption that a subject needs to be able to consciously grasp the fundamental rule of reference—and hence all the involved concepts—if she applies a concept in thought. Subjects don't need to be capable of conceptualising the fundamental rules of reference of a concept in order to employ it in thought. After all, a child may believe *That apple is red* without being able to think about concepts like that of thinking or of perception. But both of these are involved in the fundamental rules of reference for the concepts employed by the believing child.

It seems that the requirement of tacit knowledge leads to problems similar to the ones encountered by the linguistic approach. Before going on to discuss these issues in more detail, let me quickly recapitulate the conceptual approach. I will use the following general characterisation of what I take to be the essential part of the approach:

CONCEPTUAL APPROACH TO DE SE THINKING
When a subject thinks about herself in the *de se* way, she entertains a thought that's constituted by her tacit knowledge of the fundamental rule of reference of the first person concept which she employs in her thinking.

The conceptual approach tries to capture what's characteristic about *de se* thinking via the grasp of the first person concept in its application in thinking. Since the concept is such that it always refers to the subject that employs it in thinking, the approach easily gets the semantic features right. When a subject believes *I'm tall*, she employs

the first person concept which makes sure that she herself is the intentional object of her thinking. The fundamental rule of reference of the first person concept is such that it always refers to the thinking subject. Furthermore, the concept ensures that the conditions of satisfaction depend systematically on whoever entertains the *de se* thought. Because the thinking subject is the referent of the first person concept, the satisfaction of any *de se* attitude depends on who's thinking.

Now, it's important to note that there's only one first person concept. So, two subjects thinking in the *de se* way are both employing the same concept. This is because the first person concept is individuated by its fundamental rule of reference. And that reference rule doesn't change when Alpha instead of Beta employs that concept in thinking. Moreover, the epistemic grasp of the concept always amounts to some tacit knowledge of the reference rule we've established. The knowledge might be more pronounced in some cases, but it's always knowledge of one and the same rule of reference. Hence, Alpha's belief *I'm tall* is identical to Beta's belief *I'm tall* on the conceptual level. Thinking back to our three elements of thinking, this amounts to an identity of the third element. Of course, the other two elements of thinking will differ due to the context-dependence of the first person concept. After all, Alpha is thinking about Alpha and Beta is thinking about Beta. But since the way the world is presented to us is established on the conceptual level—the third element of thinking—Alpha's and Beta's ways of thinking about the world are identical.

This is somewhat opposed to other accounts inspired by Frege. These take it that every subject has an individual first person concept (A.2.7). And this concept isn't merely applied on the basis of one's grasp of the reference rule. Rather, it's applied on the basis of one's primitive and subjective grasp of oneself. Presumably, this kind of relation we have to ourselves is individual and unique. The relation that Alpha has to herself is different from the relation that Beta has to herself. This idea has its origin in a cryptic remark by Frege where he maintains that 'everyone is presented to himself in a particular and primitive way, in which he is presented to no-one else' (Frege 1956: 298).

It's possible to take Frege as holding that there are individual senses associated with every subject's particular way of thinking about herself. That's because we might want to hold on to the idea that there's only ever one thing in the world we refer to when we apply a concept. In

other words, we might want to retain the claim that sense determines reference. When we think of something *as* the most famous female novelist of the 18th century, we're always thinking about Jane Austen in our world. And when we apply the concept of *pride*, we always refer to one and the same property of pride. However, in the context of two-dimensional approaches to *de se* thinking we discovered the need and possibility to have only one first person concept. One single concept that's capable of referring to different things in different situations. This allows that all *de se* thoughts are fundamentally of the same type. They only differ insofar as they're thought in different contexts by different subjects. And such an account easily explains why subjects draw the same rational inferences on the basis of their application of the first person concept. If they literally apply the same concept—and not a concept which is merely in the same ball park as the other concept—it's clear why they should reason similarly on that basis.

Now, this semantic theory is complemented by an important epistemic element that attempts to explain the non-semantic features of *de se* thought. Tacit knowledge of the rules of reference is required to apply a concept in thought. This addition elucidates why immunity to error through misidentification is characteristic of *de se* thinking. If a subject grasps the fundamental rules of reference, then there's no open question for her concerning the intentional object of that event of thinking. Presumably, a subject can't grasp that her use of the first person concept is such that it always refers to the thinking subject and at the same time wonder who she's thereby thinking about. This fact isn't grounded in the semantic nature of the first person concept. It's rather established in the epistemic aspect which resides in the subject's tacit knowledge of the fundamental rule of reference.

So far so good. However, I've already hinted at some problematic similarities between the linguistic and the conceptual approach. We can ask whether the regress problem doesn't peak its head out here as well. In our discussion of the linguistic approach, we saw that a strange identification between the first person pronoun and its *de se*-devoid character gave us a serious headache. Assenting to a sentence reporting the character of the first-person pronoun isn't the same as assenting to the *de se* sentence. If we now look at the conceptual approach, we notice that the definition of the fundamental rule of reference of the first person concept is equally lacking what's characteristic for *de se*

thinking—an unmistakable epistemic self-reference to the thinking subject. There's nothing *de se* about the fundamental rule governing the reference of the first person concept. It merely describes under which conditions subjects necessarily refer to themselves.

In the context of the linguistic approach, I argued that the subject can reasonably ask who the speaker of (8) is despite the fact that that sentence accurately reports the character of 'I'. And since the character marks the cognitive significance of an expression, the account doesn't represent the epistemic feature of *de se* thinking accurately.

A similar defect can be identified in the context of the conceptual approach. Here, we're confronted not with the character but with the fundamental reference rule. And similarly, the *de se* role of the first person concept isn't exhausted by the description 'the subject who's thinking *this* thought', which expresses its fundamental rule of reference. A subject can reasonably believe that the subject who's thinking this thought is tall while doubting whether she herself is tall. This is because she can wonder whether she herself is the subject who's thinking this thought. Hence, mere grasp of the fundamental rule of reference of the first person concept doesn't seem to be enough to account for what's characteristic of *de se* thinking. A subject needs to be capable of self-referring in the *de se* way in order to grasp the self-referring nature of the first person concept. But because the conceptual approach puts all the explanatory weight on the grasp of the reference rule, it can't incorporate this more basic, primitive way of thinking about yourself.

This new regress problem reveals its full force when we turn to the connection between *de se* thinking and intentional action. While believing *I'm in danger* ensures that the believing subject will take the necessary actions to get out of danger, the same doesn't apply to her belief *The subject of this thought is in danger*. This is despite the fact that the latter expresses the fundamental rule of reference of the former and is thus what individuates the relevant first person concept.

What is it for a subject to believe *I'm in danger* on the conceptual approach? Most importantly, it involves the application of the first person concept which refers to the thinking subject in all cases. This semantic fact is coupled with the requirement that the subject should possess tacit knowledge of this semantic rule. The subject needs to know in some—to be established—way that her use of the first person concept always refers to the thinking subject. But how does that help?

A subject might be fully aware that her use of the first person concept in believing *I'm in danger* is governed by the fundamental rule of reference. She might even have full knowledge of that rule. Nonetheless, she wouldn't be moved to action on the basis of that semantic knowledge. This is because she doesn't think of herself in the *de se* way when she's aware of the fact that subjects *in general* necessarily self-refer when they employ the first person concept. We can make this argument stronger and more precise by approaching it from the point of view of rationality. It might be rational for a subject to judge *The subject of this thought is in danger but I shouldn't run away*. This is because her knowledge of the semantics of the description 'the subject of this thought' is distinct from her knowledge that the description refers to *herself*. Similarly, the knowledge of the semantics of the first person concept is distinct from the kind of *de se* grasp we're after.

In fact, the theory of concepts we're using, which works in the background of the conceptual approach, comes back to bite us. It holds that the objects of thought are presented to us in thinking via the concepts we use. But it doesn't seem that subjects are presented to themselves in *de se* thinking via knowledge of some complex fundamental rule of reference. It's rather the opposite. We think about ourselves first in the *de se* way and only later possibly grasp the fundamental rule of reference governing our use of the first-person concept. In other words, our acquisition of the first person concept depends on our prior ability to think in the *de se* way. And therefore, not every instance of *de se* thinking is governed by the application of the first person concept.

Let me give a quick argument for this claim. For the first person concept to play a grounding rule for *de se* thinking, it has to be necessarily self-referring because every instance of *de se* thinking is about the thinking subject. Any subject that employs a concept in thinking has to tacitly know the referential nature of the concept. Hence, a subject employing the first person concept needs to know that the first person concept is necessarily self-referring. The knowledge of the necessary self-referring nature of the first person concept is *de se* because the subject needs to know that she's thereby referring to herself. Therefore, a subject employing the first person concept needs *de se* knowledge in order to apply the first person concept.

This argument supports our earlier rebuttal of the conceptual approach. Rather than thinking about ourselves in the *de se* way on the

basis of employing the first person concept, that conceptual capability depends on some prior *de se* knowledge. Otherwise, a subject couldn't grasp the self-referring nature of the concept. So, something more epistemically basic on the non-conceptual level has to support the role that the first person concept plays. Something without which the conceptual approach doesn't even get off the ground. This doesn't mean that any account of *de se* thinking requires knowledge of the necessity of self-reference. This would be much too strong. But, if we think that the ability to think about oneself is guided by the application of the first person concept, then such knowledge is crucial because it's implied by the necessary grasp of the relevant concept.

In all fairness, Peacocke is well aware of this inevitable path to some more primitive conception of *de se* thinking. He anchors the tacit knowledge underlying our grasp of the first person concept in what he calls the non-conceptual first person *notion*. He explains that 'this connection with knowledge is explained by the relations between the first person concept and the more primitive nonconceptual *de se* notion' (Peacocke 2014: 86). In this way, he permits that we need to distinguish the cognitive role of the *de se* element from the semantic content of *de se* thoughts. It's one thing to identify the two-dimensional conceptual content of the belief *I'm tall*. It's quite another to explain how a subject thereby comes to think about herself in the *de se* way. While our conceptual approach does fairly well on the semantic part, it overlooks the primitiveness of *de se* thinking that governs its cognitive role.

Even with this proviso concerning Peacocke's own account—which goes beyond what we've called the conceptual approach by including an important *de se* element that's non-conceptual—one can't refrain from wondering why the conceptual approach should be especially well equipped to explain the basic way of thinking about oneself that's characteristic of *de se* thinking. Arguably, not all thinking is thinking in concepts. And because some forms of *de se* thinking seem to be cases of a primitive and basic way of thinking, there's a strong possibility that they potentially also reside in the non-conceptual realm. After all, a subject doesn't need a first person concept in order to believe that she's in pain or that her food is to the left—both instances of *de se* thinking. But if there's non-conceptual *de se* thinking, then the conceptual approach is fundamentally misguided. It misses a large chunk of what needs an explanation.

So, it seems that the focus on the first person concept is only viable in the highly rational and sophisticated thinking that's characteristic of projects like the one by Descartes. You might object that I take it for granted that there's non-conceptual *de se* thinking. And that I can't invoke this assumption against the conceptual approach. And you would be right. The fact that the conceptual approach requires an underlying non-conceptual *de se* way of thinking about oneself weighs stronger though. Not only does the conceptual approach miss a potentially primitive form of non-conceptual *de se* thinking. I also argued that without such a more primitive ability, it can't even do the job on the conceptual front. Without it, a subject's tacit knowledge of the first person concept can't get off the ground because she couldn't understand what it is for something to be self-referring in the first place. So, the first person concept derives its *de se* powers from an underlying capability of subjects which allows them to primitively think about themselves in the *de se* way. The conceptual approach is based on the grasp of the reference rule. And this grasp is only complete if it's already *de se* in nature. A subject can only understand that her use of the first person concept refers to herself if she has some prior grasp of what it means to refer and think about oneself in the *de se* way. This shows that the non-conceptual way of *de se* thinking is more fundamental than the use of the first person concept. And hence, the conceptual approach—as a whole—fails to account for the basic kind of *de se* thinking that we want to understand.

We've seen now that the conceptual approach does fairly well in a small area of typically highly rational *de se* judgements. Here, it explains how subjects come to think of themselves in the characteristically self-referential way when they employ the first person concept. But the discussion above identified two main problems for such an account that attempts to explain *de se* thinking via the use of the first person concept. The weaker first problem is that the account is too narrow by excluding many *de se* thoughts which are most likely not from the conceptual realm. If subjects can think non-conceptually in the *de se* way, then the conceptual approach doesn't provide an explanation of what's going on in these cases. The second more serious problem is that even in the case of conceptual *de se* thinking, we need an anchor that's located on a more primitive and basic level. Otherwise, the concept can't play its epistemic role aimed at characterising a special

way of thinking about an object. The reference rule alone doesn't tell us that we're actually thinking about ourselves. Hence, the conceptual approach is neither sufficient for our purposes because it overlooks important areas of *de se* thinking which need to be accounted for. Nor is it necessary since it has to rely on some more basic way of *de se* thinking in order to play the cognitive role it's supposed to.

2.3 FUNCTIONING PROPERLY

It seems that our excursion into the two-dimensional world hasn't been particularly crowned with success so far. The strategies we've discussed got the semantics of *de se* thinking right but failed on the epistemic front. But let's not get discouraged yet, because there's one final two-dimensional theory I want to look into before drawing a verdict. And what's interesting is that this approach doesn't start from the semantic two-dimensional framework we're all familiar with right now. It rather starts from the cognitive significance of our thoughts by asking about the role that mental states play in our reasoning and motivation. That certainly sounds promising because it tackles the epistemic features upfront. The idea is to commence from the epistemic and psychological role that *de se* thinking plays and try to account for the semantic features in a second step.

Why is this alternative worth our consideration? In our discussion of the conceptual approach, we've encountered a need to distinguish the cognitive role of *de se* thinking from its semantic content. It's one thing to identify the conditions of satisfaction and the intentional object of our thoughts about ourselves and it's quite another thing to explain how this kind of thinking has the peculiar epistemic and cognitive features it has. The traditional two-dimensional approaches sought to account for both of these things in one go. Now, we want to look at another option which attempts to slightly disentangle these two and deal with them individually.

Thinking about oneself in the *de se* way is special because it has a very peculiar cognitive significance. We saw this when we compared Alpha's belief *I'm tall* with her belief *Alpha is tall*. The two mental states have the same conditions of satisfaction—they're both true if Alpha is tall. And at the same time, one and the same state of affairs is given to Alpha in two very different ways. She will reason differently on the

basis of either belief. And she might be motivated to act in distinct ways. That's because the cognitive role of *de se* thinking is more intimately tied to our own action and motivation. Furthermore, this role remains the same for every subject thinking *I'm tall* despite the fact that the conditions of satisfaction change dramatically. It seems that all *de se* beliefs of the same type have the same functional and cognitive effect on a subject.

The difference in cognitive significance between mental states is at the forefront of the *functional* approach to *de se* thinking. It characterises a thought on the basis of how it's functionally related to other thoughts and actions of ours. For instance: Which conclusions can a subject draw on the basis of entertaining a given belief? What actions is she motivated to perform on that basis? What other beliefs and desires is she inclined to form as a result of that mental state? If we think about *de se* thoughts in this context, the similarities they have when entertained by different subjects in different situations immediately strike us. The belief *I'm tall* has a certain cognitive significance that stays roughly the same—independently of who entertains it. Similarly, a different *de se* belief like *I'm scared* has the same cognitive role for all thinking subjects.

This leads to an attempt to individuate our thoughts through their cognitive function. The background idea is this: If two beliefs produce the same 'output' on the basis of the same 'input', they're identical. And if two beliefs produce different outputs starting from the same input, they aren't the same—independently of whether they have the same semantic features or not. A belief like *I'm tall* is of the former kind. We get roughly the same output in all believing subjects—provided the other inputs are the same. If you and I are otherwise significantly similar, we're going to act similarly on the basis of believing that. For instance, we'll both reason that we need to duck or that we can reach the top cupboard. A mere semantic focus on that belief doesn't directly approach this similarity. It merely cares about the conditions of satisfaction, context dependence, and similar things without looking at the cognitive effect of our mental states. And in the end, Alpha's believing *I'm tall* has quite different conditions of satisfaction than Beta's similar belief. Instead of looking at the variable conditions of satisfaction, the functional approach instead shines a light on what stays the same whenever a subject entertains a certain *de se* belief. And because *de se*

beliefs have quite a different impact on our minds than *de dicto* beliefs about ourselves, they should be distinguished accordingly.

The general two-dimensional strategy involves an important distinction. On the one hand, there's a dimension that's concerned with contextual features of thought and speech. And on the other hand, we have a referential or truth-conditional dimension. For instance, the character of an expression is located in the former domain because it allows us to determine the truth conditions of an utterance in a specific context. Without the context, we wouldn't get anywhere. Or, if we look at the first person concept, we see that it appeals to the context of thinking. We only know who's referred to by a specific use of that concept once we fix the context of that particular mental state. The functional approach also distinguishes the contextual, or functional, dimension from the referential, semantic, or truth-conditional dimension. The former is concerned with the cognitive role of our thoughts while the latter tells us a thought's conditions of satisfaction. That is why the functional approach—though slightly atypical—is two-dimensional in nature.

Again, we can track the origins of the functional approach back to Frege. But it receives a quite pronounced expression in another influential paper by John Perry titled 'Frege on Demonstratives' (1977), where he discusses some consequences of Frege's distinction between sense and reference for demonstrative thoughts such as *That apple is green*. When he gets to *de se* thoughts, Perry reminds us that we need to keep in mind the way subjects think about the intentional objects of their thoughts in order to correctly understand and analyse a given mental state. More importantly, we need to look at the cognitive significance or function of a thought for a subject:

> We use senses to individuate psychological states, in explaining and predicting action. It is the sense entertained, and not the thought apprehended, that is tied to human action. When you and I entertain the sense of 'A bear is about to attack me,' we behave similarly. We both roll up in a ball and try to be as still as possible. Different thoughts apprehended, same sense entertained, same behavior. When you and I both apprehend the thought that I am about to be attacked by a bear, we behave differently.

> I roll up in a ball, you run to get help. Same thought apprehended, different sense entertained, different behavior.
>
> Perry 1977: 494

Perry's claim here is that mental states have to be characterised by the way things are presented to us in thinking and not by the intentional objects of our thoughts. This is because one and the same intentional object can be presented to us in different ways and produce different behaviour. And in such a case, we're left with quite distinct mental states because the world is presented to us in distinct ways. So far, this isn't a particularly new insight. However, Perry focuses on the neglected fact that these mental states have different connections to our own actions and our other beliefs and desires. We can take Perry's example to illustrate this. Here, we have one and the same state of affairs—one and the same proposition—which can be presented to Alpha in two different ways. Alpha can think about the proposition <Alpha, about to be attacked by a bear> either from her own first personal perspective or from an impersonal bird eye view. In the former case, she would be motivated to behave in quite different ways than in the latter case. In fact, Perry argues that we shouldn't concentrate on the mere semantic question, 'What are you thinking *about*?', in the first place because it doesn't fully reflect the different ways that things can be thought about. Rather, we should depart from the idea that we can characterise mental states merely on the basis of the proposition that we thereby entertain—or the 'thought apprehended' in Perry's terminology. It's important that we need to incorporate the way we think about things into our theory.

As mentioned, this doesn't sound particularly new yet. But Perry doesn't just repeat the Fregean insight that lead us to the distinction between sense and reference. We have to understand his point as a focus on the connection between our beliefs and desires on the one hand and our actions and behaviour on the other. There's more to the story than merely doing justice to the different perspectives that subjects have on the world. Beyond that, there are important relations between the way we think about the world and how we are motivated to interact with that world. How things are presented to us in thinking is crucial for understanding the nature and role of our mental states in the broad picture which includes our behaviour as well.

When two subjects both believe that they're about to be attacked by a bear, they normally behave similarly. Their mental states have the same functional role despite having a different semantic profile. They motivate us to do certain things and don't do others. The way things are presented to us—being associated with the Fregean sense—is part of the explanation of why we behave in certain ways.

The argument we extracted from Perry can thus be understood in the following way. We can observe a clear difference between Alpha's belief *I'm being attacked by a bear* and Beta's belief *Alpha is being attacked by a bear*. This difference manifests itself not only in the way things are presented to the believing subject. Beyond that, we find disparate kinds of behaviour that are the result of these distinct beliefs. It's true that both beliefs have the same semantic object. They both present the proposition that Alpha is being attacked by a bear to a believing subject. However, if there's a difference between the two beliefs, something has to make that difference. And this difference has to be located on the functional level because the two beliefs are semantically identical in that the same proposition is presented to us in thinking.

Let's quickly compare the functional approach with the other two previous candidates in order to carve out the dissimilarities. The linguistic approach focused on the mere semantic profile of the character of an expression. The character of an expression is a tool that gives us the right semantic value for a linguistic expression in all kinds of conversational contexts and situations. In order to do justice to some of the epistemic requirements, Kaplan imposed a certain necessary epistemic access onto the semantic story. We clearly witnessed the failure of that project. On the other hand, the conceptual approach attacked the epistemic issues more directly. It directed its attention on the way the world is presented to us in thinking by explaining that certain ways of thinking about the world coincide with the application of certain kinds of concepts. Again, I explained the problems of such an approach.

Now, the functional approach builds on the useful elements of both these approaches and adds its own twist. It remarks that mental states have a certain cognitive role. And this functional aspect needs to be at the forefront of the characterisation and individualisation of our mental lives. Perry's main observation is that two subjects that behave similarly in a similar situation should be said to have the same kind of *de se* belief. In other words, the functional focus ties in our mental lives

with our interaction with the world—after all, that's what mental states are presumably for. The central claim is that beliefs have the functional nature they have *because* they present us the world in a certain way. We can make this claim more precise in the following way: The way things are presented to us in thinking determines an intentional attitude with a specific functional role. If the world were to be presented differently, the mental state would have a different functional role. And if the belief had a different functional role, it would present things in a different way. An example helps to appreciate this interconnection. Imagine that Alpha thinks that the spider is dangerous and Beta believes that the spider is harmless. In such a case, one and the same object—the spider—is presented in two different ways to the subjects. At the same time, Alpha and Beta behave differently on the basis of these different beliefs. Beta might inspect the spider closely while Alpha tries to get as far away as possible. The functional approach can get two things out of this example. On the one hand, their distinct behaviour tells us that Alpha and Beta are in different mental states because same input should result in same output. On the other hand, the fact that they're in different mental states has to do with the fact that they think about the world in different ways (A.2.8).

Perry calls the contextual and functional dimension of our beliefs the *belief state*. The psychological state that a subject finds herself in is primarily characterised by the functional role of that mental state. It's what's common between Alpha believing *I'm tall* and Beta believing *I'm tall*. They're both in a certain state of believing of themselves that they're tall. This comes with a certain functional profile. Both subjects will reason that they have to duck or that they can reach the top cupboard. On the other hand, the semantic dimension of the belief is the *belief content*. This content gives us the state of affairs, the proposition, that the subject believes to be the case. In Alpha's case it's the proposition <Alpha, being tall>, while in Beta's case it's the proposition <Beta, being tall>. The belief content is not very different from Kaplan's content and emerges from purely semantic aspects of the belief. It's only interested in the belief's conditions of satisfaction and reference.

We're now in a position to give a general characterisation of the functional approach to *de se* thinking and then decide whether it accounts for the characteristic features of *de se* thinking. Let's proceed with the following for now:

FUNCTIONAL APPROACH TO DE SE THINKING
When a subject thinks about herself in the *de se* way, she entertains a thought that's constituted by its functional role, which is determined by the subject taking what she thinks about to be of direct importance to her own behaviour.

Admittedly, this is a quite convoluted and maybe imprecise way to characterise the functional approach. Nonetheless, let me try to illuminate it a bit through Perry's example. The intentional object—what our subject is thinking about—of Alpha's belief *A bear is about to attack me* is the state of affairs that Alpha is being attacked by a bear. The belief is *de se* because she takes that particular state of affairs to be of direct importance to her own behaviour. What does that mean exactly? In ordinary circumstances, Alpha wouldn't need any further reason beside her belief to do something about her being attacked. The *de se* belief alone is enough to motivate her to behave accordingly. She needn't think that she herself is Alpha, she needn't think that Alpha is worthy of her attention and help and so on. Her cognitive architecture is such that this kind of belief is directly relevant to her behaviour. Compare this to Beta's belief with the same intentional object. Whether Alpha is being attacked by a bear or not is only indirectly important to Beta if she believes *A bear is about to attack Alpha*. She'll only come to aid if she likes Alpha, is capable of helping, and so on. Beta's cognitive architecture requires additional reasons for that belief to produce some kind of action. Alpha's *de se* belief, on the other hand, doesn't require any additional reason to motivate her to act in some way or other. So, we might say that her belief is of 'direct importance' to her own behaviour.

The functional approach characterised our *de se* thinking via its special functional role. What's characteristic is that the things we believe about ourselves in the *de se* way are of direct importance to us. There's no gap between the intentional object and the question whether what we attribute to that object applies to ourselves or not. Subjects are immediately aware that they're self-ascribing a property. In contrast, *de dicto* thoughts about ourselves have a quite different functional role. They usually concern something distinct from us—for instance the first woman in space. And this ranges from important world news to gossip to utter bullshit; things that are usually only indirectly import-

ant to us. The disparity between the three different ways of thinking about oneself—*de se*, *de re*, and *de dicto*—is especially pronounced when we look at how these thoughts are connected to reasoning and motivation. Only *de se* thoughts serve directly as reasons for us and motivate us to do certain things. The others always require some additional *de se* element in order to be taken as concerning ourselves. It's this special intimate connection that's accentuated by the functional approach when it stresses the function of being directly important to our behaviour.

But does this account for all the characteristic features of *de se* thinking? The satisfaction of the semantic features suggests itself easily. The belief content of *I'm tall* heavily depends on who's entertaining that belief. The belief wouldn't have the functional role it has—that is, it wouldn't be the belief state it is—if the content weren't always in some sense about the believing subject. How could believing *I'm tall* be of direct importance to Alpha's own behaviour if the functional role of that belief wasn't to ensure that she's thinking about herself? After all, we can't assume that she's interested in someone else's height. However, her own height is important to her movement and her possibilities of action. The functional approach thus jumps on the bandwagon of the purely semantic two-dimensional approaches. The belief content of a *de se* belief is always about the believing subject in the now familiar context-dependent way. Hence, the conditions of satisfaction systematically depend on who's entertaining the belief. And this is because a *de se* belief wouldn't exhibit the functional role it has if it weren't of a type that's designed to ensure that the subject thinks about herself. We thus have a nice account of the semantic features of *de se* thinking.

Because the functional approach focuses so strongly on the connection between our mental lives and our behaviour, it shouldn't come as a surprise that it elegantly accounts for the fact that *de se* thinking is intimately tied to intentional action and behaviour. After all, the characterisation of the approach makes explicit reference to the fact that *de se* thoughts are taken by the subject as directly important to her own *behaviour*. This, of course, implies that other thoughts which aren't of the *de se* kind are only taken by the subject as *indirectly* important. In other words, subjects have to concatenate these other thoughts with a *de se* mental state in order to take them as their own reasons for intentional action. The notion of directness that's in play here is somewhat vague. However, it has to do with the question of whether some addi-

tional mental state or reason is required for the subject to be motivated to act on the basis of her belief or desire. Accordingly, a reason or a mental state is directly important to the subject if no other independent reason is required for the subject to be motivated to act according to that particular reason or mental state.

Let's flesh out this notion of directness a bit more. Many of our mental states only give us 'neutral' information that requires some additional state in order to move us to act in a certain way. Let me give an example to illustrate this. If a subject believes

(10) *There is water in the fridge*

she's entertaining a *de dicto* belief. But that belief isn't of direct importance to her because the semantic object of her belief—the proposition <water, being in the fridge>—doesn't stand in any kind of rational or motivational relation to her own behaviour. It's just a piece of information that might come in handy. But for all we know, our subject couldn't care less about the content of the fridge. However, if she so desires, she can use this information to achieve some of her goals or satisfy some of her needs. For instance, if she holds the additional *de se* belief *I'm thirsty*, she can link up that belief with her belief (10) to have a reason to go to the fridge and get the water she believes is in there. Without the belief (10), she wouldn't know where she can satisfy her thirst. So, while her new *de se* belief is of direct importance to her behaviour, it doesn't yet ensure successful action. However, (10) alone doesn't come equipped with the motivational power that's typical of *de se* thinking. It always requires some additional belief or mental state to be motivational. So, believing (10) is only indirectly important to a subject's behaviour because it requires an additional belief or desire to be motivating.

If we compare this with the subject's belief *I'm thirsty*, we note that this *de se* belief doesn't require further motivational input. It's directly relevant to her behaviour. Upon believing herself to be thirsty, she will think about whether she knows any places to get water or actually go out to look for a fridge or a fountain. This is because a certain state of affairs is given to the subject in such a way as to be of personal relevance—in this case her survival. The function of *de se* thinking is to present things as strictly about oneself. And this, in turn, results in it being directly important to the thinking subject. In believing *I'm*

thirsty, the subject's thirst is presented to herself as something that concerns her in a very immediate way. Moreover, the functional role of that particular belief is such that it moves a subject to find something to drink—always provided there aren't any other conflicting reasons. We can thus see nicely how the functional approach accounts for the intimate connection between *de se* thinking and intentional action.

What about the proposed immunity to error through misidentification that's typical of *de se* thinking? Here, things get a bit more tricky. It's not immediately clear why immunity should be a feature of mental states with the functional role that's indicated in our characterisation. We said that a subject entertaining a thought with immunity usually shouldn't enquire about the intentional object of her thought. If a subject believes *I'm tall*, it isn't open for discussion to her who she's thinking about. There's no room for error on the part of the subject here. I explained that one reason for this characteristic is that *de se* thinking typically doesn't involve an identification of a subject. And this is why that identification can't go wrong in such cases. How does the functional approach fare in this regard? If we look at the characterisation above, we can't detect a clear relationship between the intentional object of a thought being directly important to the behaviour of the subject entertaining that thought and the fact that this mental state is a candidate for the immunity in question. Of course, there's a way to connect the two things. For instance, one might argue that a subject can only take something to be directly relevant to herself in the way envisioned by the functional approach if that mental state is free of any identification. Were it dependent on an identificational element in thinking, the thought would only be indirectly relevant to the subject. So, one might try to argue that identification-freedom and direct importance to the subject go hand in hand.

There's a problem with this tightly knit connection though. I argued that immunity isn't a feature of the structure of a given mental state but rather a feature of the epistemic process that underlies the origin of a thought. One and the same mental state can be immune in one case and subject to error in another. The fact that a subject takes a given state of affairs as directly relevant to her own behaviour is quite independent of the epistemic question concerning the basis of her belief or judgement. Chances are that a subject takes something to be directly relevant to her actions but still wonders about the intentional

object of her thought. For instance, she might—on the basis of seeing a wound on an arm she takes to be her own—judge *I'm bleeding* and take this to be directly relevant to her own behaviour. However, due to the dependence on an identification of *some* arm as her own, the judgement doesn't exhibit immunity. In such a case, we would have direct importance without identification-freedom.

The morale of this argument is that not all *de se* thoughts are immune to error through misidentification. But, according to the functional approach, all of them are taken by the subject as directly relevant to her own behaviour. The only way to account for immunity from this perspective is to tie it in with the directness of *de se* thoughts. But this is overstraining the connection between *de se* thinking and immunity. Not all *de se* beliefs exhibit immunity, yet all of them are taken as directly relevant to the subject within the functional approach. Hence, this directness can't be taken as an explanatory reason for the immunity of some *de se* beliefs. If direct importance explained immunity, then all *de se* beliefs should be immune, but that's not true. Some weaker explanatory feature is required, but unfortunately absent from the functional picture as described here.

This plea for a more flexible connection is at the same time one of the major shortcomings of the functional approach. Merely distinguishing between the contextual and functional belief state on the one hand and the semantic belief content on the other doesn't yet tell us much about the nature of *de se* thinking. It's extremely plausible that *de se* mental states have a peculiar functional role. But that insight is utterly anaemic without a further account of what that functional role precisely consists in. In fact, the proposed connection to our behaviour and the indication of the direct importance smacks of hand-waving. We've already established that *de se* thinking is of special importance to action and behaviour. So, we've already marked the field via this specific function which distinguishes it from non-*de se* beliefs we have about ourselves. But we want to have an account of what's responsible for that feature. We don't just want to have it repeated in our definition of *de se* thinking. And unfortunately, such an explanation is absent from the picture (A.2.9).

Sure enough, Perry's distinction between belief state and belief content tells us *something* about *de se* thinking. It just doesn't tell us enough. It establishes quite clearly the peculiar functional role of our thoughts

about ourselves but it fails to give an account of what's responsible for this function. So, the final verdict for the two-dimensional strategies is unfortunately negative. None of them provides an explanation that satisfies all our theoretical demands and answers all our fundamental questions about *de se* thinking. However, the survey of these failed approaches wasn't for naught. We learned the importance of four things. First, *de se* thinking has a peculiar directness to it. This directness is especially manifest in the role *de se* thinking plays with regard to intentional action and behaviour. We literally can't help ourselves but to take our *de se* thoughts personally. Secondly, there's an important epistemic dimension present in *de se* thinking. Subjects immediately know that their *de se* thoughts concern themselves without needing to apply complex linguistic or conceptual rules. In other words, we can find a certain epistemic immediacy to our *de se* thoughts. Thirdly, and this is related to the previous point, the mode of presentation that's typical of *de se* thinking is fundamentally different from the way we think about other objects in the world. We draw different conclusions, act differently, and our thoughts are more directly about their intentional objects—ourselves. The final point will set our way for the next chapter. We need to depart from the overly intellectual way of characterising our mental states if we want to do justice to *de se* thinking. Both the linguistic and the conceptual approach started out from a focus on certain highly intellectual capacities. But neither our linguistic competences nor our conceptual abilities can account for the very basic way of thinking about oneself that's typical of *de se* thinking and grounds these more demanding other capacities. We need to turn to this primitive first-personal access in order to better understand what makes our thoughts about ourselves so special.

3

BACK TO THE PRIMITIVE

> Beyond the obvious facts that he has at some time done manual labour, that he takes snuff, that he is a Freemason, that he has been in China, and that he has done a considerable amount of writing lately, I can deduce nothing else.
>
> Arthur Conan Doyle: *Adventures of Sherlock Holmes*: 31

Whenever we think about the multitude of things in the world, we attribute properties to these things. We think of the strawberry as ripe, the day as rainy, the friend as loving and reliable. In all these cases, there's a thing we think about and a property we attribute to that thing. Of course, we can attribute many different properties to one and the same thing. And by doing that we get a pretty good picture of what kind of thing it is. In fact, many things can be singled out and individuated by providing a list of all the properties they supposedly have. In just this way, Sherlock Holmes thinks about the unknown suspect in the epigraph of this chapter. He doesn't have a name or a face yet, but Sherlock has a list of properties of which he knows that the suspect instantiates them—a fancy and more accurate way of saying that a thing has a property, because things can't literally *have* properties, for otherwise every property could only be had by one thing at a time, which isn't what we want. Once he has a list of properties, Sherlock can roam the world and look for the thing that matches his description.

This is but one example of the importance of properties to our ability of thinking about the world. They allow us to characterise the things in the world and distinguish them from one another. These characterisations can be quite general or very rigorous. To illustrate this, we

can think of Valentina Tereshkova either as a woman or as the brown-haired Russian cosmonaut born on March, 6, 1937 in Maslennikovo who, as the first woman, spent 2 days, 23 hours, and 12 minutes in space. While the former merely places her in a large group of individuals who share a property, the latter way of thinking about her ascribes a whole list of properties to an individual and sets it apart from probably all other things in our world. As a matter of fact, there's only one thing that instantiates all of these properties, whereas there are many women in our world. So, the more properties we know of something or someone, the better we know it. And this ultimately brings us closer to the thing itself. In order to unravel the mystery and ultimately identify the culprit, Sherlock Holmes and other detectives have to try to discover as many properties as possible about the suspect. So, we have another example of how properties play a crucial role in our thinking.

But what *are* properties? Here's a very short answer. Properties help us to characterise the individual things in the world. Most importantly, they're something that different things can share, such as when we say that all the apples on the tree are ripe and ready for picking. In such a case, we ascribe one and the same property to a variety of, in this case related, things. In such a way, we can group the things in the world into clusters sharing the respective properties. As already mentioned, this also means that properties are normally such that they can be instantiated by several things at the same time. But this doesn't exclude the possibility that in our world, or in all different possible worlds combined, only one thing has a particular simple or complex property. We saw that Valentina Tereshkova is the only thing in our world that 'fits' the description of being the brown-haired Russian cosmonaut born on March, 6, 1937 in Maslennikovo who, as the first woman, spent 2 days, 23 hours, and 12 minutes in space. In other words, she's the only thing that instantiates this complex property. In theory, however, there might be a different thing, maybe in a different world, who also instantiates that same property.

There are quite many different metaphysical theories about the exact nature of properties. They deal with questions such as: 'How can it be that one thing—e.g. the property of being red—can be wholly present in two different things at the same time?' or 'How can a thing actually instantiate or take part in a property?' But we need not deal with these intricate discussions and bits and bobs. What matters to us is that we

attribute properties to things in thinking, be they individual things or clusters of things. By that, we logically distribute all the things in the world into two different groups: the things that instantiate the property and the things that don't. Accordingly, the property of being red divvies up the world into a pile of red things—ripe tomatoes, human blood, some of Helen Frankenthaler's paintings, a 1986 Ferrari Testarossa—and a pile of things which aren't red—my hair, clean drinking water, some other paintings of Helen Frankenthaler.

This way of thinking should strike you as familiar, at least if you've read through chapter 2. There, I explained how concepts work in thinking. And in fact, concepts and properties are closely related. Both things are attributed to things in the world and logically divide all the things into two different piles. This is a good opportunity to make the relationship between the two notions a bit more precise. When we say that a property characterises a thing, we mean to say that the property is really instantiated in that thing out there in the world. On the other hand, concepts are tools of thinking. We apply them in thinking *about* the things out there in the world. Hence, when a concept using subject thinks that the spider is dangerous, she applies the concept of danger to the thing she thinks of as a spider. In doing that, she ascribes the property of being dangerous to that individual thing. If that thing really has that property, her application of the concept was felicitous. Otherwise, she misapplied the concept in attributing a specific property to that thing. We can also metaphorically put it this way: Properties *really* divvy up the things in the world into different piles while concepts are our tool of *thinkingly* dividing the things into piles.

The notion of a property is thus a metaphysical one. That is to say that it deals with the structure and the things that make up reality. On the other hand, the notion of a concept is epistemically flavoured. It reflects the way we think about and gain knowledge of the world around us. It's important to keep this distinction in mind. However, we can and shall talk about the ascription or attribution of properties in thinking as well. This way of describing things—and the contrast to the application of concepts—opens up our theoretical scope. It allows us to capture not only conceptual thinking but every kind of thinking in which subjects characterise the things they think about in some way.

Using this broader notion, we can say that a newborn thinks of a stick as bent by ascribing the property of being bent to the stick

but without applying either the concept of a stick or the concept of a bend in her thinking. Of course, this way of talking is still epistemically loaded since the newborn might gain some form of knowledge about the stick through her ascription of the property. Furthermore, she might be wrong in her ascription of the property. On the other hand, the stick itself either has the property of being bent or it doesn't. We don't find the epistemic possibility of false ascription of a property in reality. I will thus contrast between a property being instantiated or had by a thing and a property being ascribed by a subject to a thing. Only the latter has an epistemic touch and will prove to be a fruitful way of describing the situations we're interested in.

Naturally, we don't just ascribe properties to things in the world around us. We also attribute them to ourselves. We think of ourselves as being tall, as romantic and empathic, as late, or as drunk and in pain. These are all instances of self-ascriptions. Through these self-ascriptions, we form a picture of ourselves—whether it's an adequate picture is, of course, a different question. And every time we ascribe a property to ourselves we can be described as placing ourselves logically in one of two piles. So, when Alpha thinks that she's tall, she characterises herself as being tall. She thinks that she herself instantiates the property of being tall by ascribing that property to herself. And through that, she groups herself together with all the other tall things in the world. In the context of our expedition to learn about how we think about ourselves, these self-ascriptions are of prime interest. What's their role in *de se* thinking? Can they tell us something new which we haven't yet discovered (A.3.1)?

3.1 GODLY PROPERTIES

We saw that the Propophile way of thinking comes with a certain schema. They attempt to characterise our *de se* thoughts on the basis of a certain proposition—a possible way the world could be—that a subject entertains. I then examined three different strategies of explaining how an impersonal proposition can come to concern us in the way typical of *de se* thinking. In all these cases, we witnessed a move away from the proposition itself to the way the subject entertains that proposition. And maybe the right lesson to be learned from these failed attempts is to lay aside this Fregean relic and try things anew. Maybe, it's much

easier to understand the way we think about the world and ourselves if we simply take ourselves to ascribe properties to things.

The two famous American philosophers David Kellogg Lewis and Roderick Chisholm provide prime examples of this way of understanding things. They both want to explain how we think about ourselves in the *de se* way in terms of the properties we ascribe to ourselves. One of their central claims is that only through self-ascribing a property do we arrive at a *de se* belief. In other words, entertaining a *de se* belief amounts to the self-ascription of a property and that's it. There's no need to bring a proposition or the first person concept into play. All that's required is that subjects take themselves to have certain properties by self-ascribing them. According to this picture then, when Alpha believes *I'm tall*, she takes herself to have the property of *being tall* by self-ascribing that property. And Beta's belief *I'm tall* works in just the same way: She takes herself to have the property of *being tall* by self-ascribing that property.

How did they arrive at this theory? Well, like us, they started with the Propophile doctrine and soon realised that it can't be the whole story. Lewis in particular was interested in properties for reasons that go beyond the desire to account for *de se* thinking. He defended a theory that held that every possible world is as real as our own actual world and exists in the same way. In this theory, properties play an important role. This is because Lewis took properties to be just the set of all individual things that instantiate a particular property. He thus arrived at the claim that the property of *being red* is nothing but the pile of all things which are red, whether they're part of our world or some other far away possible world. This is a surprising claim and somewhat different from a more Platonic idea according to which a property is something abstract and outside the realm of directly accessible reality. Lewis wanted to hold on to the idea that properties are very real—just like all the possible worlds where they're instantiated. For our purposes, we don't need to deal with this debate because nothing in our endeavour hinges on the exact metaphysical nature of properties. I, for one, don't subscribe to Lewis's theory of possible worlds and properties. Nonetheless, the thesis that believing in the *de se* way amounts to self-ascribing properties is a worthwhile candidate to examine. And this is why we can just go with the Lewisian flow for the moment and throw his metaphysics over board later.

With this metaphysical theory in his pocket, he then contrasted the way properties divide logical space with the way propositions divide that same logical space. Let me explain what I mean by 'dividing logical space'. Earlier, I elucidated that propositions are generally true and false in a possible world *in toto*. That means that if the proposition <Sydney, being the capital of Australia> is false for me, then it's *ipso facto* false for every other individual in our actual possible world. Either the world is in that way or it isn't. There's no middle ground and no room for variation here. Propositions aren't true for some individuals and false for other individuals within the same possible world. Hence, if we want to draw a logical map of truth for that particular proposition within the universe of possibilities—telling us 'where' the proposition is true and where it's false—our smallest unit will be that of a possible world. We can't say that the proposition is true in one *part* of that world and false in another. This shows us that propositions can merely logically carve *around* possible worlds and divide logical space along world borders—this way of speaking should, of course, be understood metaphorically and not geometrically. Correspondingly, when I want to know what the proposition <Sydney, being the capital of Australia> logically amounts to, I will always pick out whole possible worlds where that proposition is either true or false.

Now, from this logical point of view, properties are quite different. They can be instantiated in some objects of a possible world but not in others. For instance, the property of standing is at this moment instantiated in some human beings, water bottles and lamps. At the same time, it isn't instantiated in some pens, sleeping animals, clouds and pieces of clothing. This suggests that a property can logically cut *through* a possible world and divide things within such a world. So, properties can 'zoom in' on a possible world where propositions are blind. But properties can do even more. They're sometimes instantiated in *every* object of a possible world. To give an example, let's examine the property of *inhabiting the actual world*. Quick reflection reveals that everything has that property in the actual world and nothing has it in all the non-actual possible worlds. Hence, this property picks out a whole possible world within logical space in a similar way as propositions do. These useful features of properties open up the possibility of dividing logical space in a much more fine-grained way. A specific property can be instantiated in a couple of objects in the actual world

and in a couple more in other possible worlds. So, properties are both capable of carving logical space around possible worlds, in a similar manner as propositions do, and they're able to carve *through* possible worlds and divide logical space willy-nilly.

That's certainly an interesting piece of information, but we should better put it to use somehow. And in fact, there's more to the Lewisian story. He goes on to argue that all the work propositions do can be done equally well if we just use properties. Properties are a much more flexible tool when mapping logical space and can be used in just the way that propositions function. How does that work? Well, as we just saw, there are indeed properties that carve *around* possible worlds in the same manner as propositions. They always group together whole possible worlds including all the individuals that are part of that world. We've already seen a very special such property in action: the property of inhabiting the actual world. What other kinds of properties could do that? For instance the property of *being such that Anna Magdalena Bach died in Leipzig*. This is a property that you and I—and in fact every stone, leaf, star, molecule, or Higgs boson in the actual world—have. Everything in our world has that curious property because it's a fact that Anna Magdalena Bach died in Leipzig. And this fact makes it so that everything in our world has that, admittedly very unsubstantial, property (A.3.2).

What's more important for Lewis's argument is that if we look at the proposition <Anna Magdalena Bach, having died in Leipzig> which corresponds to the fact above, we quickly realise that everything in every possible world where that proposition is true has the corresponding property *being such that Anna Magdalena Bach died in Leipzig*. So, if we peek into possible world w_{249}, where Anna Magdalena Bach also died in Leipzig, we see that everything instantiates that property just as everything has that property in the actual world. And conversely, nothing instantiates that property in all the worlds where the proposition is false. If we had a look at a possible world where she didn't die in Leipzig, nothing could possibly have that property because it's simply not true in that world that she died in Leipzig. Hence, if we overlay the logical space as divided by that proposition with the logical space as divided by that property, we get a perfect match. Isn't that great?

Unfortunately, this exercise of logical connect the dots isn't particularly impressive if we can't show that properties are *necessary* for dealing

with *de se* thinking. Of course, Lewis is well aware of this requirement and gives us an argument why we need to take properties on board and throw propositions out. In fact, he gives us both a negative argument why propositions aren't enough—something which we by now hopefully accepted—and a positive one which tells us why properties are a suitable candidate to fill the void that propositions leave. Together, they are designed to support the claim that properties are a necessary and sufficient conceptual tool to understand *de se* thoughts. Let's see how Lewis puts the negative part first.

His argument starts from the contrast between knowing something about the world and knowing something about oneself. Sometimes a subject gains knowledge about her surroundings, for instance when she learns the truth about Valentina Tereshkova's birthplace. And in other cases, a subject learns the truth about herself, such as when she knows that she herself is bleeding. Importantly, every piece of self-knowledge is at the same time a piece of world-knowledge. If Alpha knows that she's tall, then she *ipso facto* knows something about the world—namely, that Alpha is tall or that someone is tall. Of course, she might not know it in that neutral guise, but she knows something about the world nonetheless. In contrast, and trivially, not everything we know about the world is a piece of self-knowledge. If Beta knows that Valentina Tereshkova was born in Maslennikovo, she doesn't know anything about herself. Even if some piece of knowledge about the world is actually about the knowing subject, that fact might be elusive to her and result in a piece of knowledge about the world without self-knowledge. So, Gamma might know that Gamma is bleeding without at the same time knowing that she herself is bleeding. This is our old contrast between thinking and knowing *de re* and *de se*.

Now, imagine there are two omniscient goddesses. As long as we're merely equipped with propositions in our theory, we can characterise their omniscience as knowledge of every true proposition in their world. In other words, they know all the facts in their world because every true proposition corresponds to a fact. They're equipped with every possible piece of knowledge about the world. If something is a fact, then they know it and if something isn't a fact, then they don't believe it. For instance, they know that there are mountains, that there are goddesses, and that these goddesses interact with their world in various ways. Let's imagine that one of the two goddesses, Alpha,

who's known to live on the coldest mountain, throws down thunderbolts. The other, Beta, who's known to live on the tallest mountain, throws down manna.

Being omniscient, both goddesses know these facts. Taking the Propophile perspective, we would say that they both know the proposition <Alpha, living on the coldest mountain and throwing down thunderbolts> and they both know the proposition <Beta, living on the tallest mountain and throwing down manna>. So far, so good. But we can now ask the crucial question: 'Do they know which of the two goddesses they themselves are?' For instance, does Alpha know, based on her knowledge of every fact, that she herself lives on the coldest mountain? The initial hunch is to say yes—after all, they know all the facts of their world and it's a fact that Alpha lives on the coldest mountain—but Lewis provides us with a famous thought experiment which supports the negative answer:

> Consider the case of the two gods. They inhabit a certain possible world, and they know exactly which world it is. Therefore they know every proposition that is true at their world. Insofar as knowledge is a propositional attitude, they are omniscient. Still I can imagine them to suffer ignorance: neither one knows which of the two he is. They are not exactly alike. One lives on top of the tallest mountain and throws down manna; the other lives on top of the coldest mountain and throws down thunderbolts. Neither one knows whether he lives on the tallest mountain or on the coldest mountain; nor whether he throws manna or thunderbolts.
>
> Lewis 1979: 520–521

It's important here to clearly understand Lewis's claim. It's easy to imagine the two goddesses sitting on the top of their respective mountain and looking from there into their world. With this picture in our mind, it's then hard to additionally imagine them as being ignorant of their own location. But the argument isn't trying to deny that. Lewis never takes into account the specific perspective of the two goddesses. And with good reason. The kind of knowledge that propositions provide is very much independent from the specific perspective that the goddesses find themselves in. After all, propositions are in an

important way neutral to the perspectives that we take on the world. They're a representation of how the world could possibly be 'from a bird's eye view'. So, what the argument tries to show is that *if* we take their omniscience to be purely propositional—consisting of the knowledge of every true proposition *and nothing beyond that*—then they can't know where they themselves are.

How plausible is that? Lewis holds that we sometimes entertain beliefs and gain knowledge about ourselves which can't be properly represented through observing that the subject believes or knows a certain proposition to be true. Rather, in some cases, subjects believe something about themselves in virtue of ascribing a property to themselves. And this is different to believing a proposition about oneself to be true. The case of the two goddesses illustrates this possibility vividly. The knowledge that Alpha lacks is that she *herself* lives on the coldest mountain and throws down thunderbolts. She already knows that Alpha lives on the coldest mountain—corresponding to her knowledge of the proposition <Alpha, living on the coldest mountain>—but she lacks the necessary self-knowledge in the form of underlying knowledge that she *herself* lives on the coldest mountain. Coming to know that she herself lives on the coldest mountain doesn't correspond to coming to know a new proposition—it's just the proposition <Alpha, living on the coldest mountain> in a new disguise. But she already knew that proposition. Hence, no new proposition can represent this new insight.

Let me quickly summarise Lewis's argument. Propositions are such that they are true in a possible world in its totality. This is our Propophile starting point. Now, knowledge of one's own location in a world is knowledge a subject can gain. Yet, if propositions are as the Propophile envisages them, knowledge of one's own location can't be accounted for in terms of knowledge of a proposition. Hence, there's the possibility of knowing something that can't be accounted for within the Propophile story. More specifically, propositions aren't enough to give a full picture of how we think about ourselves and the world.

First off, the crucial premise of the argument is of course that the Propophile isn't able to account for a subject's knowledge of her own location. The case of the two goddesses, however, makes that premise very plausible and hence it's capable of supporting the overall argument. Alpha learning that she herself lives on the coldest mountain

isn't learning a proposition that differs from the proposition <Alpha, living on the coldest mountain>—a proposition she already knows. There's no new proposition that she could come to know which corresponds to her novel knowledge of her own location in her world. Secondly, the conclusion isn't new to us. We've reached it via a different route already. Accordingly, you might be unimpressed by this mere repetition. What's important and interesting, however, is that we arrived at the same result via a different route this time. Therefore, we have just the more reason to dismiss the Propophile idea. And moreover, any supporter of it has to overcome an additional difficulty. Furthermore, Lewis doesn't stop at this negative point but has a positive argument in store with which he aims to establish a New World Order to surpass the reign of the Propophiles. This time, it's not propositions that rule the world, but properties.

The theory that Lewis, Chisholm, and other so-called property theorists aim to defend typically has two components. First of all, *de se* beliefs are nothing but self-ascriptions of properties. If Alpha believes *I'm tall*, she doesn't believe a certain proposition in a certain first-personal way, she also doesn't necessarily apply the first person concept in her thinking. All she does is that she self-ascribes the property of being tall. Secondly, *every* belief is ultimately a *de se* belief. And more broadly, every thought is *de se* in nature. This is a quite surprising claim. Just earlier, I wrote that it's trivially true that not everything we know about the world is a piece of self-knowledge. If Alpha believes that Valentina Tereshkova was born in Maslennikovo, that doesn't strike us as particularly about herself. But now, these philosophers have the outrageous idea of claiming exactly that. How do they explain that?

It's a surprisingly simple theory because we're already equipped with a substantial theory of what a property is. So, let's try to make some sense of this curious claim. When our goddess Alpha entertains a *de se* belief like *I'm living on the coldest mountain*, she ascribes a certain property to herself. That property is the property of *living on the coldest mountain*. Everyone could self-ascribe that property and end up with a belief of the same kind. In a way, we can describe what she's doing with her belief as Alpha 'locating herself in logical space'. That means that she takes herself to be part of the set of things that live on the coldest mountain. She metaphorically takes a good look at the logical layout and determines that she's located within the group of things

that are unfortunate enough to live on the coldest mountain of their respective world. Of course, we might find other objects among that set. There might be a Yeti who lives on the coldest mountain in its world as well. But Alpha doesn't need to think about the other things that also might have that property. She only needs to think that she's part of that particular group with that particular property. Similarly, when I believe that I'm tall, I count myself as belonging to the tall population of things.

Describing Alpha's *de se* belief as locating herself in logical space is certainly very metaphorical and needs to be unravelled and demystified a bit. Taking into account what has been said about the nature of properties will help a great deal in this task. I explained that properties divide logical space into two respective groups: the things that instantiate that property and the things that don't. We can illuminate this by looking at how regular space is divided into different groups. For instance, there are different planets in our solar system and we see ourselves as Earthlings—inhabitants of that blue planet. So, regular space is divided into groups of things in a similar way as logical space is divided into sets of things that do or don't instantiate a property.

Against this metaphorical background, we can say that a subject who self-ascribes a certain property is therefore similarly identifying herself with a specific part of logical space and laying claim to her membership. Because it's possible to divide all the things in all the possible worlds into, maybe overlapping, groups corresponding to the different properties there are, we can expound that believing *I am F* amounts to laying claim to membership of the group of things which are *F*. And of course, a subject can self-ascribe several properties at once. She then identifies herself as part of the group of things which instantiate all of these properties. In this sense, then, self-ascription of a property is tantamount to locating oneself in logical space according to the properties one believes oneself to have.

That doesn't yet explain the outrageous idea that every belief is ultimately *de se* in nature. But this doesn't require a big theoretical leap anymore. Let's look at a case where our goddess entertains a classic *de re* belief like *Olympus Mons is 21'230 meters tall*. It seems obvious that this belief has nothing to do with our goddess. It's about that high mountain on planet Mars and not about herself. How can we spin this in order to get a belief that's truly about herself?

The answer lies in my earlier explanation that we can create a property for every proposition and have them both correspond to each other in logical terms. In the case of Alpha's belief, we can take advantage of the logical equivalence between the property *being such that Olympus Mons is 21'230 meters tall* and the proposition <Olympus Mons, being 21'230 meters tall>. Both divide logical space in the exact same way to produce two identical groups. Everything that has that property will be a part of a world where the proposition is true and nothing can have that property in a world where the proposition is false. The respective areas in logical space that are grouped together by these two distinct things are perfectly identical. Once this is accepted, it's easy to just transcribe her *de re* belief into her self-ascribing that property and we've got a *de se* belief. To put it differently, in *de re* believing that Olympus Mons is 21'230 meters tall, she's identifying herself as a part of the group of things that has the property of being such that Olympus Mons is 21'230 meters tall by self-ascribing that property. And this is nothing but a *de se* belief.

It now becomes even more obvious how important properties seem to be for our thinking about the world. In the theories of Lewis and Chisholm, they constitute the cornerstones of how subjects are capable of being in mental states. While the Propophiles held that thinking is constituted by a subject entertaining a proposition in a certain way, the property theorists tell us that thinking is nothing but the self-ascription of properties of various kinds. There are three reasons why such a move is deemed worthwhile. First of all, there are some mental states—*de se* ones in particular—which are especially problematic for the Propophile way of thinking. Secondly, properties are much more liberal in dividing up logical space and can therefore correspond to many more ways that subjects think about the world. Importantly, they can do the logical job that propositions do just as well; they're a perfect substitute. Thirdly and finally, thinking in the *de se* way is easily explained by holding that subjects self-ascribe properties when they think about themselves (A.3.3).

3.2 ASCRIBING IT TO YOURSELF

With this preliminary grip on the idea that *de se* thinking amounts to the self-ascription of properties, we now want to take a closer look at

how this way of describing things stands in relation to the characteristic features of thinking in the *de se* way. The upshot of Lewis's arguments is that propositions can't do the job we want them to do, but that properties are perfect for it. The resulting idea is that whenever we have a *de se* belief, we have a subject that self-ascribes some property or other—and not a subject who believes some proposition. This is supposed to give an account of how we think about ourselves in the *de se* way. In other words, it's an attempt to yield an explanation of the nature of *de se* thinking.

While the idea of self-ascription certainly has some immediate appeal and sounds rather inconspicuous, it also requires a bit more elaboration in order to be capable of providing a satisfactory account of *de se* thinking. Let's start by examining the following *de se* belief as a case study for our theory:

(11) *I'm happy.*

A subject that entertains the belief (11) believes that she herself is happy. According to the property theory, this amounts to the subject self-ascribing the property of *being happy*. Now, in Lewis's terms, 'we identify ourselves as members of subpopulations' (Lewis 1979: 519) when we self-ascribe a certain property. And because Lewis takes a property to be a set, or subpopulation, of all things with a certain quality, the property of being happy is nothing but the set of all happy things in all possible worlds. Hence, our believing subject takes herself to be a member of the group of happy things, as opposed to the sad things, in her self-ascription of that property. We could also say that she identifies herself with the happy things.

Two things are of relevance in this context. First, the self-ascription doesn't have to be appropriate. A subject can potentially self-ascribe any property she wants without there being any grounds for her to believe that she really has that property. Nothing is stopping you from believing that you're flying right now while you're sitting in your chair reading this book. You simply fail to self-ascribe a property that you actually have. We often go wrong and the theory leaves room for that. Moreover, subjects sometimes self-ascribe a property without entertaining the corresponding *de se* belief or forming the relevant judgement. For instance, I might imagine myself hitting a wonderful down the line backhand winner. But, I do this in the mode of imagination.

Hence, the self-ascription is done only imaginatively. Since these cases would have to be discussed more seriously within an account of imagination, they won't occupy us further here.

Secondly, subjects can either self-ascribe a property or they can ascribe it to other things in the world. Accordingly, Alpha can, above entertaining the *de se* belief (11), believe that Beta is happy by ascribing the property of being happy to Beta. But such a belief is crucially different from her *de se* belief, which is constituted by her self-ascribing the property of being happy. Believing that Beta is happy is a classical instance of a *de re* belief and needs to be contrasted sharply from the *de se* belief (11). The reductionist programme of Lewis and Chisholm doesn't change or hide that fact. Even if all beliefs are ultimately logically reduced to the *de se*, we need to make room for the different types of thinking about objects and ourselves on an epistemic and semantic level. And, indeed, it's seemingly easy to distinguish the two beliefs from each other. Following Lewis, we could say that believing *de re* that Beta is happy amounts to self-ascribing the property of *being such that Beta is happy*. If we believe the property theory, we get a new and different *de se* belief that we could express in the following way:

(12) *I'm such that Beta is happy.*

I hope it's obvious enough that there's a big difference between believing (11) and (12). We can start by noticing that the two properties which are self-ascribed are quite different. They aren't different in the way that the property of *being red* is different from the property of *being happy*. There's more to it. The two function logically in very different ways. The property that's self-ascribed in (11) is such that it can pick out individuals willy-nilly. Any individual in any possible world can possess it. What the property leaves us with is a pile of things with borders that run laterally to and across the borders of possible worlds. In contrast, the property that's self-ascribed in (12) is such that it can only pick out very specific groups of individuals. More precisely, it either picks out all the individuals of a given possible world or none of them—it corresponds to our good old proposition. A thing can only be such that Beta is happy if she inhabits a world where it's a fact that Beta is happy. Conversely, everything in a world where that fact obtains has the relevant property. Hence, propositions and properties can be mapped onto each other.

Now, one important duty of our theory of choice certainly is to distinguish accurately between thinking *de dicto*, *de re*, and *de se*. Most crucially, we want to clearly isolate cases of thinking in the *de se* way. So, how do we distinguish between the *de re* belief *Beta is happy* and the *de se* belief (11)? The difference between the two properties that are self-ascribed in (11) and (12) seems to be up to that task. Correspondingly, despite the fact that Lewis and Chisholm treat both of these beliefs as ultimately *de se* in nature—owed to the fact that they're both self-ascriptions of a property—they're nonetheless distinct. They're self-ascriptions of quite *different* kinds of properties. One is the property of being happy, the other is the property of being such that Beta is happy. We're then left with a typical *de se* belief about oneself in the case of (11) and a *de re* belief about Beta translated into a special *de se* belief for (12).

And we might add that the two properties aren't merely contingently distinct. They don't just happen to pick out different piles of things. Of course, believing that I'm happy is different from believing that I'm tall. I self-ascribe very different properties in these two cases since they correspond to different portions of logical space. But that's not the difference at play when we're looking at the properties involved in (11) and (12). After all, we're after a possibility to distinguish clearly between *de se* and *de re* beliefs. Here, we're faced with two properties that aren't just accidentally distinct. Rather, the property of *being happy* and the property of *being such that Beta is happy* are different by their *logical nature*. While the former individuates on the level of individual things—trees, dark matter, office supplies, or polar bears—the latter individuates on the level of possible worlds.

Let me explain this difference. Simple properties such as *being happy* or *standing* are of a kind that picks out individuals and forms groups which cut across possible worlds. And because they pick out individuals, a subject can locate herself in logical space on the level of individuals on the basis of self-ascribing them. We could metaphorically paraphrase this by saying that a subject who believes (11) is identifying herself with a certain kind of individual: a happy one. Accordingly, self-ascribing such a simple property amounts to a 'real' *de se* belief.

On the other hand, there are more complex and special properties such as *being such that Beta is happy* or *inhabiting the actual world* or *being such that Anna Magdalena Bach died in Leipzig*. These properties

don't pick out individuals that form groups which cut across possible worlds. Rather, they always pick out either all the things or none of them in a given possible world. We could say that these properties have a lower resolution than the simple properties above because they aren't as finely grained. They can't pick out individual things but only groups of things which correspond to propositions. As a result, a subject can only locate herself in logical space on the level of possible worlds by self-ascribing such a property. Again, we could put this metaphorically by explaining that a subject who believes (12) is identifying herself with a certain kind of possible world: a containing-a-happy-Beta one. Accordingly, self-ascribing such a complex property amounts to a *de re* belief—a 'false' *de se* belief.

Having identified a logical difference between these two kinds of properties, we could now say that we've clearly marked the difference between thinking *de re* and thinking *de se*. The former consists of the self-ascription of a property that individuates on the level of possible worlds while the latter consists of the self-ascription of a property that individuates on the level of individuals.

That sounds fair and square, but there's an unfortunate oversight which will teach us an important lesson. Remember that the belief (12) was introduced as the Lewisian paraphrase of the *de re* belief *Beta is happy*. But we should refrain from assuming that the subject literally thinks that *she herself* has the property of being such that Beta is happy when she believes that Beta is happy. What she does is she ascribes a property to some *other* thing—namely, Beta. The paraphrase in (12) is merely a logical translation with which Lewis and Friends explain why thinking *de re* that Beta is happy is nothing but a self-ascription of the property of being such that Beta is happy. In this way, they can unify everything under one common banner and justify their reductionist programme: every belief is a *de se* belief because every belief is a self-ascription of a property and locates a subject in logical space.

What, now, if a subject actually entertains the real *de se* belief (12)? In other words, what if Alpha is really thinking about herself and really wants to self-ascribe that weird property? How would the property theory account for that *de se* belief? Well, obviously, the subject self-ascribes a certain property. What property could it be? The obvious, and really the only, candidate is the property of *being such that Beta is happy*—the very same property that Lewis thinks is being self-ascribed

in the *de re* belief *Beta is happy*. But if that's the right way to go about it, we have two distinct beliefs which are analysed identically. That's an unfortunate consequence which should be avoided.

Contrary to earlier, there's no way to distinguish between the two properties which are self-ascribed in this new case. They're obviously identical and no philosophical spin will get us out of that trouble. The proposal of distinguishing between different kinds of properties isn't available in this case. We can't claim that one and the same property can have a different nature depending on the way it's self-ascribed. That would severely beg the question because the difference in the kind of belief is supposed to be accounted for in terms of what kind of property is self-ascribed.

Let me quickly summarise the argument behind this serious problem and offer a solution afterwards. We—including Lewis, Chisholm, and Friends—want to distinguish between thinking *de se* and thinking *de re*. However, we've now reduced all thinking to *de se* thinking understood as self-ascription of properties. This again muddles the distinction between thinking in the *de re* and the *de se* way. Our weapon of choice against this problem is to distinguish between two kinds of properties: properties that pick out individuals and properties that pick out possible worlds—or rather: properties that always pick out all or no individuals within a possible world. When a subject self-ascribes the former, she thinks in the *de se* way, and if she self-ascribes the latter, she's in a *de re* mental state. However, sometimes subjects entertain *de se* beliefs like (12) where they literally self-ascribe the latter properties. Hence, there are cases where a *de re* belief and a *de se* belief have to be analysed as self-ascription of one and the same property. And therefore, the self-ascribed property isn't enough to distinguish all cases of thinking *de se* from all cases of thinking *de re* (A.3.4).

The only way to save the property theory from the repercussions of this argument is to shift the focus away from the property that's being self-ascribed to the act of self-ascription that's involved. The source of this shift is the following observation: It's easy to comprehend *de se* thinking as the self-ascription of properties. When a subject thinks (11), she just ascribes happiness to herself. But when Alpha thinks that Beta is happy, she doesn't really ascribe a property to herself. She ascribes a property to some other thing. Now, it's possible to philosophically translate this in the Lewisian way to be logically equivalent

to the self-ascription of some complex property. But that's missing the important point. The kind of ascription that's constitutive of *de se* thinking—i.e. self-ascription—is *epistemically* special. It's a way for the subject to ascribe a property to herself in the intimate way that's typical of *de se* thinking.

This insight is especially pronounced in Chisholm's own term for self-ascription: 'direct attribution'. There's something epistemically direct and immediate going on in the case of *de se* thinking. We could express this by saying that a subject attributes a property to herself directly. In contrast, when Alpha thinks *de re* that Beta is happy, she merely attributes a property to Beta in an *indirect* way. There's an epistemic intermediary involved in the ascription of a property to something other than ourselves. Chisholm puts it in the following way:

> How does one succeed in making *other* things one's intentional objects? In other words, how is it possible to refer to individuals other than oneself? For example, how do I make you my intentional object? I would say that the answer is this: I make you my object by attributing a certain property to myself. The property is one which, in some sense, singles you out and thus makes you the object of an *indirect* attribution.
>
> Chisholm 1981: 29

Without going too deep into the interpretation of Chisholm's own theory of *de se* thinking—which is similar, but not identical to the Lewisian model we're focusing on—we can easily see that he draws a sharp contrast between the epistemic relation that subjects and their intentional objects stand in when they think in the *de se* way and the one we have in cases of thinking about other things. While it's possible to attribute properties to oneself directly, we can only ever do so indirectly when other things are the intentional objects of our thoughts.

Chisholm himself emphasises the need for Alpha to self-ascribe some relation that she's standing in with regard to Beta in a case where Alpha's trying to ascribe the property of *being happy* to Beta. Because she can only attribute that property to Beta in an epistemically indirect way, she needs to think about the way in which she is related to the intentional object of her thought. In the most neutral case, she can achieve this through the self-ascription of the overarching property of

being such that p. This covers the case where the subject merely believes that it's a fact that *p* is the case. But she might stand in a different epistemic relation to Beta. Let's imagine that Beta and Alpha are talking to each other. In such a case, Alpha might think about Beta *as* the thing she's talking to. And as a result she might come to believe that Beta is happy through the self-ascription of the slightly more complex property of *being such that the thing talked to is happy*. Chisholm's point can be understood as invoking the requirement of self-ascribing the epistemic relation one stands in with regard to the intentional object *over and above* the property one wants to ascribe to that thing. Thus, we introduce a specific epistemic relation into the property theory. This supplements the idea that we can distinguish mental states only through the property that's being self-ascribed in thinking.

We now have the two required puzzle pieces in order to explain why we need to focus on the epistemic aspect of self-ascription in order to clearly mark the territory of *de se* thinking and distinguish it from *de re* thinking. First, the blind focus on the property that's being self-ascribed results in the possibility of having distinct beliefs which are analysed as self-ascriptions of one and the same property. And secondly, a glance at the epistemic difference in how we ascribe properties to the intentional object in the case of *de se* and *de re* thinking reveals that self-ascription has to come with some directness that's not present in the case of thinking about other things.

Against this background, Lewis points to the fact that there's a special *acquaintance* with the intentional object in the case of thinking about oneself in the *de se* way. When Alpha talks to Beta, she's acquainted with Beta through the fact that she can see, talk to, feel, and hear her. And we're all acquainted with Valentina Tereshkova through the fact that we can read about her on websites and in this book. But nothing comes close to the kind of acquaintance we have with ourselves. There's nothing else we're acquainted with in this intimate and direct way. Lewis explains that this is because we're *identical* with ourselves and nothing else. He then uses this as the defining epistemic relation underlying self-ascription in the case of *de se* thinking.

This, then, is the proposed version of how the property theory accounts for the distinctiveness of *de se* thinking. It's because we self-ascribe properties under the epistemic relation of identity. In other words, the guiding relation we bear to the thing we ascribe the prop-

erty of *being happy* to in (11) is the fact that we're identical to the intentional object of the belief. We're thus left with the following general account of *de se* thinking:

THE PROPERTY THEORY OF DE SE THINKING
Whenever a subject thinks about herself in the *de se* way, she ascribes a property to herself under the epistemic relation of identity.

How do we determine the success of that account? Most importantly, it should be capable of doing justice to the five characteristic features of *de se* thinking, which were identified in chapter 1. Particularly, it needs to provide satisfactory answers to the following questions: Why are *de se* mental states always about the thinking subject? Why do the satisfaction conditions of *de se* thoughts depend in a systematic way on the thinking subject? Why are some of our *de se* mental states immune to error through misidentification? How can *de se* beliefs provide a foundation for self-knowledge? And finally: Why are *de se* thoughts so essential for intentional action and behaviour?

The first two questions are easily answered. The thinking subject can only be identical to herself. She can't stand in that relation to some other thing. Therefore, *de se* thinking will always be about the thinking subject if that subject ascribes a property to herself under the relation of identity. There's just no room for error here. The intentional object of an ascription to the thing that's identical to oneself is always the thinking subject herself.

Similarly, the conditions of satisfaction of a *de se* belief will depend systematically on who's self-ascribing the property. More precisely, who the thinking subject is identical to is determined by who the thinking subject is. The relation of identity is a reflexive relation and naturally takes into account who's ascribing a property under that relation. Hence, we can only determine the conditions of satisfaction if we know who the thinking subject is. Once this is known, we know which thing is identical to the thinking subject: she herself. Let me illustrate this: If a subject believes that she's tall, she ascribes the property of *being tall* to herself under identity. And if we want to know the conditions of satisfaction of that *de se* belief, we need to know who the identical thing is. And that, of course, will depend on who's thinking. If Alpha believes that she's tall, it's Alpha who's identical to the thinking subject. Thus, it will be Alpha who needs to be tall in order for her

belief to be true. And the same logic applies to the case where Beta believes herself to be tall.

With the first two features out of our way, we now come to an explanation of immunity. I argued that the basis for immunity to error through misidentification lies in the fact that a subject needn't identify herself with a certain intentional object of her thought. Rather, *de se* mental states can be free of identification. For instance, this is the case when Alpha thinks *I'm in pain* on the basis of experiencing the pain in her arm directly. There's no need for her to identify a thing in the world as the thing in pain before ascribing pain to that thing. The way the pain is given to her doesn't leave room for the question concerning the subject of that pain. It seems that the way the property theory handles *de se* thinking is well equipped to account for the possibility of thinking about oneself without the need for identification. There's no first person concept or pronoun involved. Rather, it's a simple case of self-ascribing a property. And because we can't self-ascribe a property to some other thing, it's only natural that the subject needn't identify herself beforehand. Self-ascription is always ascription to oneself. Hence, the question of misidentification doesn't emerge. This fact is also reflected in the conceptual proximity between the concept of direct attribution we found in Chisholm and the concept of epistemic directness we observed in the elucidation of immunity to error through misidentification.

If you're now skeptical about the success of the property theory in that regard, your skepticism will be rewarded shortly when I'll go on to discuss some problems for Lewis's account which will lead to its partial dismissal. There's a real question about the epistemic addendum we had to provide. If a subject ascribes a property to herself under the relation of identity, isn't that a case of identification? And isn't then all *de se* thinking poisoned with an identification element? If you're worried about these questions, I ask you to stay patient until the next section where I'll discuss this objection in more detail.

What about the possibility to serve as a foundation for self-knowledge? A first issue is the earlier confession that subjects can potentially self-ascribe any property they like. The property theory remains mute on the epistemic grounds on which a subject ascribes a property to herself. So, there's no special guarantee that self-ascribing a property leads to self-knowledge. However, this issue sounds more problematic

than it is. In fact, the connection between *de se* thinking and self-knowledge we want to explain isn't very tight or even constitutive. We don't expect every case of *de se* thinking to constitute a case of knowledge. The only requirement is that *de se* thinking can potentially, under the right circumstances, result in cases of self-knowledge which might be of a distinctive kind.

The obvious question for the property theory is then: Which aspect of the theory matches the peculiarity of self-knowledge? Or, to put it differently: What part of the property theory explains why self-knowledge is potentially special? Let's take the case where a subject self-ascribes the property of *being happy* on the basis of her feeling good. In such a case, she's entertaining the *de se* belief (11). Earlier, I explained that one possible explanation of the peculiarity of self-knowledge is through alluding to the distinctive epistemic path which leads to a subject knowing her own mind. Because subjects have a direct and immediate access to their own mental states, they're epistemically well positioned and authoritative with respect to the contents of their minds. In contrast, a subject ascribing happiness to some other subject has to rely on some indirect epistemic path to the mental state she's trying to ascribe. She might have to ask the other subject or observe her behaviour. The epistemic directness present in the case of observing one's own mind produces the possibility of self-knowledge. Accordingly, if a subject gains a true belief about herself via some direct epistemic route, she's in a distinctive state of self-knowledge.

Now, the property theory is very much capable of representing this contrast. The solution is to point to the different forms of acquaintance which underlie *de se* and *de re* beliefs. I explained that Alpha's *de se* belief (11) is formed on the basis of a self-ascription under the acquaintance relation of identity. In such a case, we can say that the self-ascription is direct and immediate. By contrast, Alpha's mock *de se* belief (12) is formed on the basis of a self-ascription under some indirect relation of acquaintance. She has to observe Beta, read about her, or ask her. Distinctive self-knowledge is therefore formed if a subject correctly ascribes a property to herself under the relation of identity from an epistemically direct source. Hence, if the self-ascription happens under the right direct circumstance and is correct, the subject can achieve self-knowledge. We thus see that the property theory has the potential to explain the possibility of self-knowledge through the dis-

tinct epistemic relation that underlies *de se* thinking. You might think that this leads to the false claim that all true *de se* beliefs are forms of distinctive self-knowledge. This isn't necessarily the case. In those situations where the property is self-ascribed on the basis of some intermediary information—such as seeing oneself in the mirror—the path of the self-ascription, but not the self-ascription itself, is indirect and thus doesn't give rise to a distinctive kind of self-knowledge.

Last but not least, let's discuss the relation between *de se* thoughts and intentional action. The property theory can neatly account for this connection. In order for a subject to be motivated to act on a reason, she has to take that reason as her own reason. For instance, Alpha's hunger is a reason for her to get something to eat. However, she will only act on that reason if she takes it as her own reason. What better way than to self-ascribe it? Self-ascription achieves exactly what's required in that a subject ascribes the hunger to the thing that's identical to herself. In this way then, thinking in the *de se* way amounts to taking oneself to have certain properties. And having these properties might present a reason for the subject to behave in a certain way. Alpha takes her hunger to be a reason for herself to get something to eat because she self-ascribes that property and thereby takes herself to be hungry. And that again is directly relevant to her intentional action.

This survey shows that the property theory seems well equipped to account for the five characteristic features of *de se* thinking—as long as we bracket a proviso that popped up in our discussion of immunity. It gives a picture of the distinctiveness of thinking in the *de se* way and it explains how *de se* thoughts are always about the thinking subject, how immunity is possible, how it might be a basis of self-knowledge, and the essential connection to intentional action. But typically, there are certain problems which have been glossed over for the sake of intelligibility and easy understanding. Let's turn to these issues now and see whether they require a substantial revision of the theory or not.

3.3 PROBLEMS ABOUND

As we've seen, the relation of identity plays a crucial role in the property theory. It's needed to distinguish the kind of self-ascription which results in proper *de se* beliefs from the self-ascription that's involved in *de re* thinking. Only in the case of a *de se* belief like (11) is the sub-

ject ascribing a property to herself under the relation of identity. The subject exploits the unique relation she has to herself in her self-ascription. In contrast, a *de re* belief results in the self-ascription of a property under some other acquaintance relation to the intentional object of the belief. In this case, the subject exploits some other relation of acquaintance—such as seeing or hearing—in her self-ascription. So, identity plays the crucial role of isolating *de se* thinking from other kinds of thinking. But how do we have to understand the claim that proper self-ascription is ascription under the relation of identity?

Let me start with the observation that identity is usually understood as a metaphysical relation. It's a relation that obtains between things in the world. It isn't a relation that materialises on the basis of our ability to think about the world. As a metaphysical relation, it's independent of our perspective on the world. Two things are identical or not solely based on how they really are. Accordingly, we might say that two things are identical when they have the same properties or when they're literally one and the same thing. In the stronger, latter sense, which is relevant here, I'm identical to myself and nothing else. And this fact about me is wholly independent of any subject thinking about me or knowing anything about me.

Now, what does it mean for a subject to ascribe a property to herself 'under the relation of identity'? We might be tempted to take the metaphysical sense of identity as constitutive of *de se* thinking. However, this approach immediately gets us into trouble because a subject can ascribe a property to the thing that's *actually*—and therefore metaphysically—identical to herself without thereby entertaining a *de se* belief. We've already witnessed this possibility many times. Alpha can believe of Alpha that she's happy without believing that she herself is happy simply by entertaining the *de re* belief *Alpha is happy*. In such a case, she thinks of the intentional object which is identical to her that she's happy. But she's thereby not entertaining a *de se* belief yet. That would require some grasp of the fact that she herself is Alpha. Hence, the mere fact that the subject ascribes a property to the thing which is identical to the thinking subject doesn't yield a *de se* belief. We need something more substantial.

Sometimes, the relation of identity gets a certain epistemic touch. Think about the case where you believe that Wonder Woman is the superheroine wielding the Lasso of Truth. In this instance, you think

that the person whose name is 'Wonder Woman' and the person who's a superheroine and wields the Lasso of Truth are identical. In other words, you think of them as being identical. Now, this identity that you establish might obtain on a metaphysical level or it might not. What's important is that identity is an *epistemic* relation in this case because you think of Wonder Woman *as* the superheroine wielding the Lasso of Truth. And it's this sense of identity that's employed in the definition of the property theory of *de se* thinking. A subject who's entertaining a *de se* belief ascribes a property to herself *as* the thing identical to her. In other words, she ascribes something to the thing she takes herself to be identical with.

We can elucidate this epistemic process by coming back to the idea of indirect attribution we found in Chisholm. There, we saw that subjects are perfectly capable of ascribing a property to things other than themselves. For instance, Alpha can think that Beta is happy by ascribing happiness to Beta. If we paraphrase this according to the scheme of the property theory, we want to say that Alpha's ascription involves some relation of acquaintance to the thing she's thinking about. For instance, when Alpha is talking to Beta, the relevant relation of acquaintance is that of 'talking to'. Accordingly, she thinks of Beta *as* the thing being talked to and Alpha takes the relation of 'talking to' to obtain between herself and Beta. This point can be generalised: Whenever a subject ascribes a property to an individual, she ascribes that property under some epistemic relation of acquaintance which she takes herself to stand in with regard to the intentional object of her thought.

Now, if we transfer this model to the case where the subject ascribes a property under the relation of identity, we simply get a special instance of the general idea that a subject always ascribes a property to a thing under some acquaintance relation or other. In the reductionist spirit of Lewis, Chisholm, and Friends, thinking in the *de se* way is just the special case where a subject takes herself to be acquainted with the intentional object of her thought in the identity way—as opposed to the talking to way or the reading about in the newspaper way. Hence, we get the following picture for Alpha's belief (11): When Alpha thinks that she herself is happy, she thinks of Alpha as the thing identical to herself. In such a case, we would say that she ascribes happiness under the epistemic relation of identity which she takes to obtain between herself and Alpha. In other words, she takes the rela-

tion of identity to hold between herself and the intentional object of her thought. But presumably and unfortunately, such an account has troubles explaining the possibility of immunity to error through misidentification (A.3.5).

Why's that? Think back to Evans's claim that immunity arises from a lack of identification in the mental state. In those cases where we're presented with immunity, the mental state only involves an ascription of a property but no identification of the intentional object. For instance, when Alpha judges that her own legs are crossed based on her proprioception, she needn't base this belief on a prior identification of these legs *as her own*. Such a belief, then, is immune to error through misidentification because the intentional object isn't identified. Rather, the belief is such that only one intentional object comes into consideration—Alpha herself. Similarly, in the case of believing *I'm in pain*, there's no possibility of misidentifying because there's no identification involved. The feeling of pain which serves as a basis for the subject's belief is such that it can only arise from herself.

On the property theory of *de se* thinking elaborated above, an identification of an intentional object is constitutively present. It maintains that every instance of thinking in the *de se* way is an instance of self-ascribing a property. And conversely, every instance of self-ascribing a property is a case of ascribing a property to an intentional object under a certain epistemic relation. In the case of thinking about oneself in the *de se* way, the epistemic relation under consideration is that the subject takes herself to be identical to the intentional object of her thought. Accordingly, when Alpha thinks that her own legs are crossed, she takes the relation of identity to obtain between herself and Alpha. We saw that we have to understand the aspect of the subject 'taking the relation of identity to obtain' in an epistemic way, which involves the identification of a subject *as* the one being identical to oneself. In an analogue way, Alpha identifies Beta *as* the one being talked to when she ascribes happiness to her under the epistemic relation of 'talking to'. Hence, a subject taking the relation of identity to obtain is tantamount to identifying the intentional object of her thinking. And this obviously annihilates the possibility of immunity for *de se* thinking.

By characterising *de se* thinking via some relation of identity that the subject takes to obtain, we introduce a source of error. While the metaphysical relation of identity can't fail to obtain, the epistemic rela-

tion is certainly fallible. I can err about the fact that Wonder Woman is identical to the superheroine wielding the Lasso of Truth. And Alpha can also be mistaken about the presumed identity between herself and the intentional object of her thought, as in the case of self-ascribing a property on the basis of seeing what she takes to be herself in the mirror. This is very bad news for the property theory. Where there is the possibility of error, we lose the desired immunity. If we understand Alpha's self-ascription of happiness as an ascription of happiness to the thing she takes herself to be identical with, we can't explain the possibility of immunity to error through misidentification because every case of self-ascription—even the most epistemically basic ones—would be subject to error.

Let me summarise this argument from immunity. One of the key premises and characteristic features of *de se* thinking was the concession that some *de se* mental states exhibit immunity. I then adopted Evans's account of immunity according to which this feature arises from freedom of identification. Now, the property theory delivers us the idea that thinking in the *de se* way amounts to ascribing a property under the epistemic relation of identity. This amounts to a subject ascribing a property to the thing she takes herself to be identical with. I argued that such an ascription involves identification of an intentional object. Hence, it precludes the possibility of immunity. Therefore, no *de se* mental states are immune to error through misidentification in the property theory. This final conclusion stands in direct contradiction with one of our characteristic features and has to result in the dismissal of the property theory as it stands.

The upshot of this rather complex discussion about identity and acquaintance is the following. In whichever way we understand the Lewisian supplement 'under the relation of identity'—either metaphysical or epistemic—it doesn't generate the desired result. If it's merely understood as a metaphysical relation, it's too weak to accentuate the special way of thinking in the *de se* way. And if it's taken as a more demanding epistemic relation where the subject takes herself to be identical to the intentional object of her thinking, it's too strong to be capable of allowing for the possibility of immunity. Unfortunately, we can't just erase that problematic supplement because I argued that it's necessary to clearly distinguish *de se* thinking from other kinds of thinking. Without such an epistemic supplement, all thinking would

be 'proper' *de se* thinking and thus anything special about our ability to think about ourselves would be blurred or lost.

Maybe the description of these problems doesn't move you to give up the idea that some form of identity has to play a role in setting apart *de se* thinking. Maybe you think that it's neither the *de facto* metaphysical identity which is relevant nor a more epistemic *assumption* of identity but rather something mysterious in between the two—strong enough to distinguish *de se* thinking but weak enough to still account for immunity. However, it's doubtful whether such a middle ground can be established. It's certainly true that identity plays a crucial role in *de se* thinking because of the semantic fact that the intentional object of *de se* thoughts is necessarily identical with the thinking subject. But we have to tread warily to not overrate its explanatory and distinguishing capabilities. Rather, it might be a consequence of the ability to think in the *de se* way without being in any way illuminating.

In order to convince you that identity has to be constitutive of *de se* thinking without playing an epistemic role, we might draw a parallel to cases where a subject is reasoning in the first person. In these cases, identity plays a crucial role without coming in an epistemic garment. Let's imagine Alpha believing *I'm happy* and also believing *I'm in love*. She can now reason from these two *de se* beliefs and arrive at a new belief *I'm happy and in love*. Why is that inference valid? John Campbell argues in his *Past, Space, and Self* (1994) that it's because Alpha 'trades on the identity' between the 'I' in her belief *I'm happy* and the 'I' in her belief *I'm in love* in her reasoning:

> This inference is valid as it stands. It is not enthymematic; there is simply no need for an identity premise. (…) What we now want to understand is how one manages to keep track of oneself through the course of the inference, how one's grasp of the first person entitles one to conclude that one and the same thing is both *F* and *G*.
>
> Campbell 1994: 84

Why is it that these kinds of first person inferences are valid without requiring a suppressed identity premise? First off, the inference can't depend on Alpha's assumption that the first instance of 'I' refers to the same thing as the second instance of 'I'. Why not? Suppose the inference was based on such an implicit identity premise *'I' in 'I'm happy'*

and 'I' in 'I'm in love' refer to the same thing. It certainly looks as if such a premise would make the inference formally valid. But this is only a deceiving appearance. Couldn't the subject then inquire again about the identity of 'I' in this new premise and the previous occurrences of the first person? This can't be necessary because it would land us in an infinite regress. If there's no guarantee of referential stability of the first person between *I'm happy* and *I'm in love*, then there can't be any such guarantee for additional identity premises involving the first person. Rather, the subject has to take it for granted that her employment of the first person is referentially constant in her reasoning.

This kind of 'taking for granted' isn't supposed to be an epistemic leap of faith. It rather tells us something important about the epistemic role of the identity relation in first person thinking. A subject doesn't have to empirically keep track of herself in order to self-ascribe properties and reason in the first person. She needn't ask herself which thing is identical to herself in order to believe that she herself is happy. For if it were necessary for the subject to keep track of herself in this empirical way, she could always be wrong about which thing she's identical to. Nothing guarantees that subjects are better at keeping track of themselves than at keeping track of the contents of their purses. Both are empirical and fallible ways of thinking about an object. And if this error-prone way of keeping track of oneself were constitutive of first person thinking, our *de se* thinking couldn't get off the ground. We could never be certain that we're thinking about ourselves. Campbell argues accordingly:

> If there really were such a thing as keeping track of oneself through the course of the inference, then it ought to be possible for the inference to go wrong because of a failure to keep track. But there is no such possibility, so long as the first person is in use.
>
> Campbell 1994: 91

These considerations are supposed to tell us that the inference from *I'm happy* and *I'm in love* to *I'm happy and in love* is valid because it doesn't involve any identification of an intentional object. If such an identification were involved, the subject might be mistaken in thinking that the first occurrence of the first person is the same as the second. But she can't be mistaken about that, for otherwise reasoning in the

first person wouldn't be possible. Hence, in order to be a valid form of inference, her reasoning has to trade on the identity that's given to her in the way she thinks about herself—the *de se* way—without epistemically taking that identity to hold. One way to make sense of this is to conclude that identity doesn't play the illuminating and distinguishing role it's supposed to play in the property theory of *de se* thinking. Rather, identity is one aspect among others which has to be present in order for *de se* thinking to take place. Moreover, our ability to think about ourselves depends on the fact that subjects can trade on that already present identity when they reason in the first person.

The connection between *de se* thinking and intentional action gives us another vivid example of why the identity relation underlying self-ascription can't be taken as epistemic but has to be traded on instead. The way the property theory accounts for the peculiarity of *de se* thinking requires inclusion of the Lewisian identity supplement. And this identity supplement has to be understood in an epistemic sense; thus thwarting the possibility of intentional action because it results in an infinite regress of required *de se* thoughts—which culminates in the impossibility of *de se* thinking.

Let's start with a simple example we're familiar with by now. The fact that Alpha is hungry is a reason for her to get something to eat. However, I argued that only a *de se* belief like *I'm hungry* can motivate her to act on that very reason. The simple obtaining of the impersonal fact that Alpha is hungry alone doesn't suffice. She has to take the reason as a reason for herself and this requires a *de se* mental state. Now, suppose that Lewis is right that thinking in the *de se* way is tantamount to ascribing a property to oneself under the relation of identity. And this again is nothing but a subject taking herself to be identical to the intentional object of her thought. How then does Alpha come to take the reason to get something to eat as her own reason?

Presumably, her belief *I'm hungry* is sufficient to motivate her to act. It's *de se* in nature and results in Alpha taking a reason as her own reason culminating in her acting accordingly. If we now analyse her belief further, we see that it's an ascription of the property of *being hungry* to a thing she takes herself to be identical with. How would we break this down? Well, if the identity relation carries some epistemic weight, it would involve some kind of ascription of the form *being such that the thing I'm identical with is hungry*. And this again involves

some kind of identity statement of the form *I'm identical to a* where *a* is whatever intentional object the subject thinks she's identical with. It's clear that such an ascription or mental state is just a new *de se* state in disguise. As such, it would require a similar analysis involving a further *de se* state, and so on.

According to the interpretation of the property theory developed in this chapter, a subject would be incapable of taking a reason as her own simply because that would involve infinitely many ascriptions of properties to oneself 'under the relation of identity'. Every single self-ascription would entail an identity ascription in the *de se* form, again requiring further self-ascriptions. And this isn't a mere psychological impossibility. More seriously, it results in the fact that *de se* thinking, and therefore the possibility of being motivated to act in a certain way, could never get off the ground (A.3.6).

We saw that the elaboration of the property theory required an analysis of what Lewis and Friends mean by a subject ascribing a property 'under the relation of identity'. We explored this idea and followed it in several possible directions, only to see that problems abound on every road, ultimately leading us to a dead end. There are two options here. Either we give up on the idea that *de se* thinking can be analysed as the self-ascription of properties, or we try to make the most of our little trip to Property Land. I propose to take the second option in compliance with the motto: 'We are experiencing trouble on every side, but are not crushed; we are perplexed, but not driven to despair' (2 Cor 4: 8–9). So, let's resolve our perplexity.

3.4 PRIMITIVE RELATIONS

It might seem to you that the discussion so far revealed an utter failure of the property theory of *de se* thinking. But such despair isn't necessary yet. Rather, I now want to argue that our discussion uncovered only an apparent failure. In fact, we've learned a crucial lesson on our way to this seeming impasse. I will now work out the exact details of that lesson, which will ultimately lead us to the claim that we have to understand self-ascription as an epistemically *primitive* relation which isn't dependent on any further and more basic epistemic constituents. In this way, we preserve much of the original idea of the property theory while abandoning some problematic aspects.

We started our exposition of Lewis's property theory with the innocuous observation that subjects are capable of ascribing properties to themselves as well as to other things—like Sherlock Holmes in his quest to find the criminal. I then explored the development of this insight into a fully fledged theory about the nature of thinking in the *de se* way. Along the way, we made contact with the idea that *de se* thinking is characterised by involving the self-ascription of properties. As such, the idea was still half-baked and required some further elaboration. This lead to the final proposal that self-ascription is nothing but the ascription of a property to oneself under the relation of identity.

Unfortunately, this proposal proved problematic because it involved the requirement for the subject to epistemically assume that the relation of identity holds between herself and the intentional object of her thought. I then argued that this can't work in order to build a foundation of *de se* thinking. Crucially, it involves an unwanted identification of an intentional object—something incompatible with the required possibility of immunity—and leads to a destructive infinite regress regarding the motivation for intentional action. We can now ask: At what point in this journey did we take a wrong turn? My proposal is that we should go back and have another look at the notion of self-ascription.

The impossibility of making sense of the role of the epistemic relation of identity tells us something about the epistemic process underlying self-ascription and *de se* thinking respectively. We've seen that it can't depend on the identification of an intentional object. Rather, it has to be more epistemically basic. It simply isn't possible for a subject to only gain basic knowledge of her own properties through knowledge of the properties of some other thing in conjunction with knowledge about the identity between herself and that thing. If a subject is capable of self-ascribing properties at all, there needs to be a point at which this self-ascription doesn't depend on such a two step epistemic process. Accordingly, if we want to understand and preserve the possibility of substantial *de se* thinking, we need to completely abandon the idea that subjects acquire basic knowledge about themselves on the basis of knowing something *else*. In other words, self-ascription can't be the result of a more basic epistemic process—like the ascription to oneself under the relation of identity. Rather, we have to take self-ascription as *epistemically primitive*.

We can give a simple argument for this claim before we elucidate it further. We start with two possibilities: Either self-ascription is primitive, or it depends on some other more basic epistemic process. Now, if self-ascription depends on some other epistemic process, it involves the individuation and identification of an intentional object because the subject has to know that her thought is about *herself*. Presumably, every such case of individuation and identification of an intentional object with oneself involves self-ascription. After all, something has to be thought of as identical to *oneself*. Hence, any account of non-primitive self-ascription is circular in depending on self-ascription itself. Therefore, self-ascription has to be primitive.

The argument as it stands is rough and simple indeed. But our discussion of the characteristic features of *de se* thinking, the different two-dimensional approaches, and the property theory has provided us with good reasons which support the individual steps. Most importantly, it painted a very convincing picture according to which our capability to think about ourselves is something epistemically basic which can't be reduced in any way. It rather seems that the ability to think about oneself in the *de se* way is at the beginning of our epistemic journey—an insight which is especially pronounced in Descartes's *Cogito* argument that we encountered.

Hence, the claim that self-ascription is primitive shouldn't come as a surprising result of our inquiry. Given what we've learned so far about the characteristics of thinking about oneself in the *de se* way, it's rather something that comes very natural without the air of mystery. In this vein, Shen-yi Liao explains why self-ascription has to be primitive:

> This mystery is to be expected given the main lesson from the problem of essential indexicals: the *de se* cannot be reduced to the *de dicto*. There is something special about learning who *oneself* is that cannot be captured in learning about what features one possesses, even if that list of features is exhaustive. There seems to be a fundamental conceptual distinction between ascribing properties to *oneself* and ascribing properties to an individual possessing a unique and exhaustive list of non-trivial properties.
>
> Liao 2012: 314

Let me explain his reasoning in the following way. John Perry originally presented us with the problem of the essential indexical, where he argued that the expression 'I'—being such that it always refers to the speaking subject—can't be reduced to or exchanged by any other linguistic expression and still retain its characteristic semantic, epistemic and psychological features. In other words, the first-person pronoun is an *essential* indexical, which results in the fact that we can't reduce *de se* sentences like 'I'm tall' to *de dicto* sentences like 'The speaker of this sentence is tall'.

In our little journey, we witnessed that this seemingly purely linguistic phenomenon is mirrored in the mental realm. Thinking in the *de se* way can't be broken down to thinking in some special *de re* or *de dicto* way. It's something completely different to think about yourself in the *de se* way than it is to think about something in virtue of it satisfying some description—as in the case of *de dicto* thinking—or in virtue of it being a specific thing we're directly acquainted with—as in the case of *de re* thinking. Our capability of thinking about ourselves isn't dependent on some other epistemic ability that we have to activate first. Rather, it's epistemically basic. But what does that mean?

In the case of primitive self-ascription, it's the ascription of a property to oneself that takes place on an epistemically basic level. The fact that one ascribes a property to oneself is thus independent of any prior epistemic achievement. Rather, self-ascription is epistemically foundational. We can illuminate this by comparing the epistemic process of primitive self-ascription with the process that underlies the ascription of a property to some other object. In the latter case, our subject first has to individuate and identify the intentional object of her thinking in some way or other. She might do this demonstratively via an epistemic relation of acquaintance. This happens in the case of Alpha thinking about Beta in virtue of talking to her. Or the subject might do this via some description she entertains as in the case of Alpha thinking about Valentina Tereshkova in virtue of the description 'the first woman in space'. In both these cases, there's an epistemic step *prior* to any ascription of a property to the intentional object. The subject has to achieve some kind of identification of an object in order to successfully ascribe a property in thought to that identified object. As such, any ascription of this kind can't be epistemically primitive because it isn't independent of some other epistemic achievement.

In Liao's terms, we can say that in the case of a subject ascribing a property to some other thing, the subject thinks about an intentional object in virtue of some feature that the thing possesses—be that the feature of being talked to or being the first woman in space. Accordingly, in cases of thinking *de re* or *de dicto*, a subject has to have some epistemic means of identifying the intentional object of her thinking. She might achieve this through picking out the intentional object via some uniquely identifying relation she takes to obtain to it, or via entertaining some description which she presumes to uniquely fit the desired object. The result in both these cases is an *indirect* epistemic process in which a subject ascribes a property to the intentional object.

But this kind of indirect epistemic route is neither necessary nor sufficient to think about oneself in the *de se* way. Alpha doesn't have to know all the features she possesses in order to be capable of thinking about herself. In fact, no subject is able to know all her features. And likewise, she doesn't have to know which identifying description is satisfied by her alone to think that she herself is tall. So, it can't be necessary to take this epistemic stepping stone we find in the cases of thinking *de dicto* or *de re*. Furthermore, even if such an extraordinary feat were possible—as in the case of our two goddesses—that kind of extensive knowledge doesn't suffice to produce *de se* thinking. The ability to think about oneself goes beyond knowing all the facts there are to know about oneself—including all the relations that various subjects and objects stand in. As Liao says, something epistemically 'special' is happening when a subject gains *de se* knowledge of herself (A.3.7).

In this way then, thinking about oneself in the *de se* way at the same time falls short of thinking about oneself in the *de re* way and goes beyond it. It falls short of it because it doesn't require some very detailed and profound knowledge of the nature of the intentional object of the thought. We can think about ourselves in the *de se* way without knowing anything particular about ourselves. And it goes beyond it because it enables a whole new kind of knowledge about oneself. Thinking in the *de se* way allows the kind of intimate knowledge about oneself in which the subject is aware that she's thinking about herself.

But how does the idea that self-ascription is primitive salvage the property theory? That entirely depends on how much of the original thought you want to preserve. Is it necessary to hold on to Lewis's metaphysical claim that properties are sets of possible individuals? I

don't think that's necessary. A subject can self-ascribe a property independently of the metaphysical dispute concerning the nature of properties. What we're interested in is an epistemic question which has only little bearing on the metaphysical details underlying the structure of reality. After all, we can perfectly well understand what it means for a subject to ascribe redness to the apple if she thinks *The apple is red* without knowing exactly what the nature of properties is. So, the same should apply to understanding what it means for a subject to self-ascribe happiness.

A more daunting question is whether one can still subscribe to the reductionist programme of Lewis, Chisholm, and Friends. Do we still want to say that every case of thinking is a case of self-ascribing a property and hence a case of thinking in the *de se* way. I would argue that this depends entirely on the exact way you spell this out. Here's what we *don't* want to say: A subject thinking *I'm tall* self-ascribes a property in the same way as a subject self-ascribes a property when she's thinking *Beta is tall*. If we want to call the latter a case of self-ascription at all, it's necessary to distinguish the two clearly from an epistemic point of view. While the former is in some sense epistemically direct, the latter is only ever indirect. Here's what we *might* want to say instead: A subject thinking *I'm tall* is in a mental state which can be logically characterised as a subject locating herself within the group of tall things. In the same way, thinking *Beta is tall* is a mental state which can be logically characterised as a subject locating herself within the group of things which inhabit a world where Beta is tall. However, this equivalence—call it 'subsuming' the *de dicto* under the *de se* if you must—is merely located on a logical level and has no epistemic repercussions (A.3.8).

And now, here's what I *advise* you to say: A subject thinking in the *de se* way is in a mental state which is constituted by her primitively self-ascribing some property or other. This form of self-ascription is epistemically basic and direct and results in the subject taking the property to apply to herself and not some other thing. Furthermore, this self-ascription is primarily characterised epistemically, which means that it's independent of the exact semantic nature of *de se* thinking. This means that the idea might be compatible with a diverse group of theories concerning the semantic contents of intentional attitudes. You might hold that they're properties, centred worlds, horizontal or

diagonal propositions. The only thing that needs to be preserved is that such a theory incorporates the special way of thinking which is constitutive of primitive self-ascription.

More precisely, the main argument is that *de se* thinking amounts to self-ascription of *properties*. But, this is just one strategy to elucidate the peculiarity of *de se* thinking. We might trace our steps back to the two-dimensional picture from chapter 2 and develop a new theory with propositions in play. It might be more complicated and less direct than the strategy of saying that *de se* thinking is self-ascription of properties. But as long as it embeds the central insight that *de se* thinking involves some epistemically primitive relation to oneself, it gets the most important component right.

Now that we've established the importance of some notion of primitive self-ascription, what's left to do is to give a more detailed account of what that amounts to. This will be the ambitious task of the next and final chapter.

4

ORIGINS

> Listen, Robert, going to another country doesn't make any difference. I've tried all that. You can't get away from yourself by moving from one place to another. There's nothing to that.
>
> Ernest Hemingway: *The Sun Also Rises*: 9–10

Primitive self-ascription lies at the basis of *de se* thinking. It is the key that unlocks the ability of subjects to think about themselves in that peculiar and direct way. But what *is* primitive self-ascription? I've argued that this kind of self-ascription is the way a subject is capable of ascribing any property to herself directly. In that sense, primitive self-ascription isn't dependent on some other prior epistemic achievement. Rather, it's the rock bottom way of coming to grasp that you're thinking about yourself.

On our journey to give an explanation of the peculiarity of *de se* thoughts, we realised that there's a specific necessity for an epistemically basic relation which grounds a subject's ability to think about herself. No other, more demanding, account can escape this requirement because every subject has to have some grasp of the fact that she's thinking about herself in employing the first person concept, the word 'I', or some other more explicit tool of thinking about oneself. Without such a grasp, she wouldn't understand that she's thinking or talking about herself. Hence, whatever primitive self-ascription amounts to, it has to be fit to play this crucial foundational role.

The necessity of primitive self-ascription for an account of *de se* thinking thus established, we want to know more about it. So far, I've merely provided a formulaic schema without any kind of content that

could give us an insight into its nature. At the moment, we might have a good grasp on what the job of primitive self-ascription is, and what its general epistemic layout has to look like. But now we have to paint the picture in more vibrant colours and fill in the gaps. Only then do we achieve a more comprehensive understanding of the underlying constitutive features of a subject's ability to primitively self-ascribe properties and *ipso facto* to think in the *de se* way.

Let me start this endeavour by providing an example of a case of *de se* thinking. Imagine a painter who's about to paint her first stroke. She feels the movement of her left arm as she leans in closer to the canvas. She wants to paint something in the top right and her arm moves appropriately without her having to somehow interact with it first. She needn't command her arm to move in a specific way. She also doesn't have to think of her arm explicitly as *her* arm. Rather, she interacts with the world *through* her arm. And similarly, she thinks about the glass of water which is given to her as being to the left and as containing drinkable water.

Why are these cases of interest to us? All these things obtain their place, as perceived by the subject, through the way the painter thinks about the world surrounding her. The canvas, the brush, the glass, the water; they all bear a certain relation to the body through which she interacts with them. Because of this layout, she can come to believe *My arm is bent* simply by primitively self-ascribing the property of having a bent arm. In that self-ascription she doesn't need to think about the arm as an object out there in the world which she then takes to be part of herself. Rather, she feels the bend in the arm with which she holds the brush and is able to paint on the canvas. The arm—and other parts of her body—is an integral part of herself and doesn't need to be identified first.

But why, you ask, is it that the arm is given to her as an integral part of herself? Why can she self-ascribe the property of having a bent arm in this primitive way? To answer this question, we need to take a bit of a detour and take a closer look at the way subjects think about themselves and the world around them. Here's an illustration of one apparent way to describe the situation: Subjects live in a world surrounded by objects with which they can interact in different ways. Our painter sees the canvas in front of her and thinks of it as something which can be transformed into a painting through her movement and

suitable tools. She imagines and intends to paint a picture of what she sees and feels when she takes a walk from her house into her garden and looks at the industrial ruins from a certain point of view. She possibly understands that what she's painting is but one way of looking into the world around her—after all, it's but a small excerpt of all the possible pictures she could paint of her world.

But there's more: Our painter isn't a mere perceiving machine with a particular point of view. Rather, she's capable of moving through her world, not just perceiving but also interacting with the things she encounters there. She sees the ruins now in front of her, now in the back. She takes the brush now with her right hand and now with her left. She knows that she has to paint the ruins in a specific way in order to reproduce the way she experienced them in her garden. This two-way interaction—our subject's ability to both perceive and manipulate the world—takes place from a certain point in her world: its centre. Everything's related to that centre in some way or other by being perceived by or by originating from it. So, in believing *My arm is bent* our subject comes to ascribe a property to the centre of her world: herself.

Of course, our painter can abstract from that centre and figuratively 'move away' from it. For instance, she can imagine the world from an impersonal point of view, abstracting from what's specific about her own perspective. And she can think about things which aren't directly accessible from her centred world view. She might know how the back of her house looks without seeing it directly, she might know how to grab a handle which is still out of reach. Importantly, she can imagine the way something looks like from a point of view which isn't currently her own. Finally, she might even turn her 'gaze' inward and think about the centre of her world from a point of view she can't literally take up. In other words, she might explicitly think about herself.

All these exercises of changing the point of view in her imagination or by walking around don't change the fact that our painter experiences her world from its centre and not from some other point of view. Correspondingly, every attempt of 'getting away' from that centre amounts to mere pretending. The epigraph of this chapter from Hemingway's novel beautifully expresses this impossibility. There's no way for our painter to literally get away from the centre of her world to some other place in her world. She can move around, but she can't get away from herself. Why? Because every one of her moves simul-

taneously moves the centre of her world with it. Where she moves, the centre of her world moves. And so, there's nothing to the attempt of getting away from yourself by moving from one place to another. Even in the imagination we can't get away from ourselves by imagining *being* somewhere or someone else.

This way of describing the situation of a subject in her world is reminiscent of a set of remarks by Wittgenstein in his *Tractatus Logico-Philosophicus* (1922). In these passages, he briefly discusses the nature of solipsism—the idea that only one's own mind can be known to exist. If solipsism is true, then the nonexistence of everything else is compatible with our experience of supposedly existing things. In this context, Wittgenstein provides us with some fascinating claims on the epistemic and metaphysical nature of subjects:

> 5.632 The subject does not belong to the world: rather, it is a limit of the world.
>
> 5.633 Where *in* the world is a metaphysical subject to be found?
>
> You will say that this is exactly like the case of the eye and the visual field. But really you do *not* see the eye.
>
> And nothing *in the visual field* allows you to infer that it is seen by an eye.
>
> Wittgenstein 1961

We can either read these seemingly apodictic statements in a metaphysical sense or an epistemological one. Since our project is of the latter kind, let's proceed thusly. What does it mean that the subject doesn't belong to the world but rather constitutes its limit? One way to make sense of this claim is that subjects don't strictly belong to the world of objects from which they can gain information. They constitute the innermost limit from which the world is perceived and established. Hence, a subject can think about the body that's hers or the social persona that she is, but she can't think about *herself* as an object *in* the world she experiences.

This is surely a bit puzzling. Isn't it weird to say that the world around us is established and *constituted* by the subject's specific perspective? If we talk about the metaphysical world—utterly independ-

ent of a subject thinking about it—then such a claim sounds mad indeed. But we're reading the remark in an epistemological sense now: The world *as found by the subject* is constituted by her way of experiencing it—for instance with things being left or right *of her*—and what she takes to be possibilities of interacting with it. This also implies that the way she can gain knowledge of the world depends on her way of being in and experiencing the world. Put this way, the claim doesn't sound mad at all.

But this isn't the only thing that's puzzling in Wittgenstein's remarks. Doesn't the idea that a subject can't literally think about *herself* as an object clash with the fundamental ability to think in the *de se* way? Don't we want to say that subjects are capable of thinking about themselves? After all, what else is *de se* thinking than the ability to think about oneself? To dissolve this puzzle, Wittgenstein's analogy to the visual field is illuminating. We can understand Wittgenstein here as holding that we, as subjects, are like an eye that enables us to make visual reference to all kinds of things relative to it. But we can never make direct visual reference to the eye itself. In other words, we never directly *see* the eye. Even if we see our reflection in the mirror, the visual experience of our own eye is only derivative. We don't see the eye where it's actually located—after all, our eyes aren't in the mirror. Rather, we see it from a strange and atypical outside view. The analogy is then supposed to show that the nature of subjects is quite similar. They aren't part of the world they experience in the same way as other things are part of the world. Rather, they're the origins of that world and everything in the experienced world is 'coordinated with it' (Wittgenstein 1961: 5.64). The subject as the origin of her world can't experience the centre—the world's innermost limit—in the same way as other objects she finds around her. In other words: Subjects are the limit of the world, but thinking about the limit of the world is fundamentally different from thinking about objects within the experienced world (A.4.1).

How does that link up with our notion of primitive self-ascription and the previous illustration of the painter? First of all, the way our painter experiences the world has to be described from a specific first personal perspective. But this perspective goes beyond the fact that we can represent and picture a scene from various points of view, each producing different ways of experiencing one and the same thing. The

world is given to our painter in a very specific way—with herself at its centre and with various ways of interacting with it being constitutive of it. In other words, the world isn't given to the painter as an allocentric, third personal, 'objective' space. Also, the world isn't just a perspectival space, constituted by particular representational 'lines of sight'. Rather, our subject finds herself in *egocentric* space. This space doesn't involve the subject itself in the way it involves other objects, but it's nonetheless intimately connected to the subject's particular way of being in the world. The experiencing and acting subject at first isn't an object in the world. Rather, it's what establishes how things in the world are given to her. Only in a second step might the subject turn her mental 'gaze' inward and think of herself as an object in the world.

Why is the notion of egocentric space connected to primitive self-ascription? The answer lies in the fact that the subject is very clearly located within egocentric space. If there's anything like an egocentric space, then the subject is always at its centre and can't fail to be there. This sounds trivial, but stands in stark contrast to the location of a subject in allocentric or perspectival space. Here, a subject can be located almost anywhere. When we look at the concept of egocentric space, we find a clear conceptual and epistemic priority of this one central place over any other place because everything is defined with reference to that specific point. Accordingly, the glass is *to the left* of the book in virtue of standing in a certain relation to the centre of egocentric space. In contrast, allocentric space knows no such conceptual priority of a place—every place stands on equal footing.

Secondly, both our painter and the subject in Wittgenstein's passages are capable of interacting with the world *from* its centre. Our subject isn't a mere geometrical point at the centre of egocentric space. Rather, she's an acting thing with various ways of manipulating the world around her. If she wants to move to a point in space, she may use her legs in a way which doesn't normally depend on some prior ability. Accordingly, she needn't first 'find' her legs in the world and then command them to move in a specific way. Quite the opposite. Her legs are an essential part of the centre of egocentric space and are given to her in an intimate and direct way. Similarly, her judging the wall to be high enough to jump over depends on her knowledge and grasp of the athletic capabilities of her own body. To say that a specific part of the body is part of the centre is to do justice to the idea that the

centre isn't a geometrical point. We'll see that it's determined by the various abilities of perceiving and acting upon the world that subjects have. As such, the subject's legs are only one among other parts that enable certain interactions.

This is relevant to the notion of primitive self-ascription insofar as the body through which our subject experiences and interacts with the world has some epistemic and pragmatic priority. This is because that acting body is the primary means of being in contact with the world. It's in virtue of the fact that she can move her left arm directly and feel the resistance of the canvas through her arm that she takes her arm to belong to the centre of her world. Similarly, the fact that she's capable of moving her legs directly and feeling the sensation of the soles of her feet makes it so that she takes them to be an integral part of herself. I'll call this phenomenological and behavioural centre the *origin* of egocentric space and argue that a subject is constituted by it.

Given this picture, we can connect this in the following way to *de se* thinking: When a subject thinks about herself in the *de se* way, she ascribes a certain property to herself. And this depends on an ascription of a property to the origin of egocentric space insofar as the *de se* nature is established therein. In the case of our painter thinking *My arm is bent*, we can explain her primitive self-ascription of the property of having a bent arm in the following way: It amounts to ascribing that property to the origin of her egocentric space—which leads to her taking herself to have that property. According to this idea, we can trace back essential primitive self-ascription to ascription to the origin.

How does this help? The fact that even in complicated and fancy science fiction scenarios there's only ever one origin for a given subject is a suitable explanation of the peculiarity of *de se* thinking. Hence, there's no question of identification involved with regard to the origin and the subject can only ascribe properties to one singular origin: herself. In this way, the origin is given to the subject in an epistemically direct way. It doesn't require any prior epistemic achievement to know which thing the origin of one's egocentric space is. As such, it can play the required role to provide a basis of *de se* thinking.

Of course, this is a programmatic claim which will need more elaboration and justification. This sets the primary task for this last chapter. We want to understand what makes primitive self-ascription so special in terms of ascription to the origin of egocentric space. And we want

to understand how that explains *de se* thinking. Before moving to business, we may summarise these introductory remarks about the relation between primitive self-ascription and the way subjects experience and interact with the world around them by repeating the tentative characterisation of primitive self-ascription from chapter 1:

> PRIMITIVE SELF-ASCRIPTION
> A subject primitively self-ascribes a property *P* iff she ascribes *P* to the origin of her egocentric space.

In this characterisation, the connections to egocentric space and its origin are made explicit and provide the keys to illuminating the crucial notion of primitive self-ascription. Once we have a clear enough account of these illuminating notions, we should be able to understand how primitive self-ascription can serve as the required fundamental capability of subjects which grounds their ability to think in the *de se* way. Moreover, they should provide a guide to account for the characteristic features of *de se* thinking.

Accordingly, it's now necessary to crystallise these notions and extract what has been implicitly alluded to in our short introductory and speculative discussion of the painter and Wittgenstein. Only through clearly shaping these notions will we be able to properly assess their merit. In this context, we also want to know why the concepts of egocentric space and its origin are necessary to understand primitive self-ascription. Hence, our task is specified and we can move to action.

4.1 FLOWING FROM THE CENTRE

I described the situation of the painter as that of a subject who finds herself at the centre of what can be called egocentric space. This kind of space was briefly contrasted with both allocentric and perspectival space. Importantly, we have to draw the distinction in terms of how a subject thinks about the world around her and not in terms of differences 'out there' in the world—e.g. by assuming the existence of two different kinds of space. As Soldati (1998: 122) argues, there's 'no egocentric space to be opposed or added to public or objective space. Rather, there is one sort of space, represented either egocentrically or allocentrically'. But what exactly is constitutive of egocentric space? What distinguishes it from these other possibilities of thinking about

the world? And why is it necessary to describe the subject's world with which she engages in this way?

Let's start with the following observation: In egocentric space, objects are given to the subject in certain terms which are relative to her. The most colourful pair of such terms is maybe *left* and *right*. On the basis of these terms, subjects can think about the objects in the world around them as being to the left or right of other things. What role do these terms play? In her experience of the world around her, our painter might think of the industrial ruins as being to the left of her garden. But this needn't be constant. Once she moves around, this way of thinking about the relation between the two objects is bound to change. After a while, the ruins might instead be to the right of her garden. Hence, it makes sense to claim that thinking about the world in terms of things being to the left or right of other things is dependent on the thinking subject engaging with the world from a certain perspective—a certain point of view that she occupies.

A first approximation to define egocentric space is to point to its perspectival nature. However, this characterisation isn't specific enough to clearly mark the territory of egocentric thinking. This is because subjects can think about space from a certain perspective without that implying that the subject takes herself to be at the centre of this world—in other words, without it constituting egocentric thinking. Hence, not every way of thinking that involves terms like *left* and *right* is egocentric. Rather, a subject can think about space in a perspectival but 'detached' way. For instance, Michelangelo's *The Creation of Adam* pictures God and Adam from a certain perspective. It shows Adam as being to the left of God and maybe located slightly below as well. Now, you can imagine yourself at the origin of the perspective which produces this particular scene—maybe by standing at just the right spot in the Sistine Chapel. But this doesn't yet allow you to interact with the objects you experience in this specific space. For, how would you have to move in order to stand next to Adam or behind God? The way things appear to the subject in a perspectival way in this particular picture doesn't inform the subject about her position relative to the things she thus experiences. A subject can take up the natural centre for such a picture—the point in space from which it is intended to be looked at—without thinking about Adam or God in an engaging way. In other words, she can think about these objects in a perspectival way

without taking a stance on her relative location. Quite the contrary, she's capable of taking an outside perspective on the painting.

Thinking in egocentric terms is thus more distinct than merely thinking in perspectival terms. I want to argue that it involves some connection to our assumed possibilities of action in the world. Why is that so? Imagine a situation where you think about some item of food in a merely perspectival way as being located in the kitchen. In this situation, you're thinking about your kitchen in merely perspectival terms. You picture the food as being in the fridge which is to the left of you once you enter the door, and in the bottom shelf once you open the fridge. Clearly, such a kind of thinking—involving such terms as *to the left* or *at the bottom*—is perspectival and thus depends on a certain point of view. We might say that you imagine your kitchen from a certain point of view when you think about it in perspectival terms. But do you know how to go about reaching the food on the basis of thinking about the objects in your world in this way? Not necessarily. This is because you might be oblivious to your spatial relation to the kitchen and thus incapable of finding the food in the world you're actually in. You simply don't know how to get to the kitchen. Only once you think about the kitchen in relation to your own current location will you know how to move in order to get something to eat from there. But this requires you to think of the kitchen in egocentric terms, locating it relative to where you think you are right now and relative to how you would move about in order to interact with it.

Let me put this point in somewhat more general terms. We want to establish that any way to think about space in terms which aren't egocentric is insufficient to coordinate the behaviour of the subject with the things she's thereby thinking about. Now, subjects come to interact with objects around them by knowing where they themselves are located with respect to these objects and what they think they can do with them. Thus, the way they think about the world has to provide them with a basis for this kind of knowledge. The claim is the following: While thinking about the world in egocentric terms indicates one's own location relative to the things in the world, perspectival or allocentric thinking makes no such reference. Thus, only egocentric thinking enables interaction with the things thus thought about. In other words, only by thinking about objects egocentrically—i.e. in terms which are relative to one's own assumed possibilities of

interaction—is a subject able to coordinate her perception of the world with her behaviour in that world (A.4.2).

The kitchen example from earlier provides us with concrete support for this claim. In this case, we witnessed that perspectival thinking—thinking that involves a privileged point of view but no reference to the behavioural centre—isn't up for the job. It's certainly true that a perspectival way of thinking about the world includes a particular point of view and thus a specific location which constitutes the centre of the world thought about in that way. However, that centre isn't necessarily the point from which the subject acts. The subject needn't take herself to be located at that particular location. She's capable of thinking about the world in perspectival terms in a totally detached way without taking up the central position.

This can be contrasted with the egocentric way of thinking which is constituted by the two-way relation between thinking about the world and engaging with it. Subjects take themselves to be the origin of their world; they're at the centre of egocentric space. And because subjects engage with the world from that particular, epistemically and conceptually privileged location, the origin isn't given to the subject in a detached and neutral way. Rather, she necessarily takes herself to be at the centre of the world thought about egocentrically. John Campbell makes a similar point in his *Past, Space, and Self* (1994) when he writes about egocentric frames of reference:

> Any animal that has the relations between perception and behavior needed to direct action at particular places, to reach for things it can see, must be capable of this egocentric spatial thinking.
>
> Campbell 1994: 5

Campbell's claim here is that subjects who can interact with the world have to be able to direct their behaviour at particular places in the world which they perceive and think about. And for this to be possible, a subject has to engage in egocentric thinking. This way of thinking puts the objects in the world in a straightforward relation to the thinking and interacting subject. In egocentric thinking, the subject thinks 'about the space from a particular point of view, as a subject at the center of one's world' (Campbell 1994: 6). This point of view is more demanding than the mere visual perspective we find

in realistic paintings. In other words, a subject who thinks about the world in an egocentric way takes herself to be located at its centre *as the subject* engaging with the world (A.4.3).

So, we've worked out the following two claims: First, an egocentric way of thinking about the world is necessary in order for a subject to be able to interact with the things she encounters in the world on the basis of thinking about them in that particular way. And secondly, the egocentric way of thinking is characterised by the subject's assumed possibilities of interaction with the objects in her world. These two claims are closely connected, but why should we accept them? Let me attempt a defence.

Thinking back to Lewis's story of the two omniscient goddesses gives us the material to support the first claim. The goddesses's omniscience can be understood as a case of thinking about the world in an absolutely impersonal way and additionally getting the facts right. In other words, their perfect knowledge doesn't involve any particular privileged point of view. They 'see' the world from no particular point of view without there being a centre of the world. We may say that their knowledge consists in a bird's eye view on the world with the facts 'laid bare in front of them'. In this way, they know that Valentina Tereshkova was the first woman in space. They know that Olympus Mons is 21'230 meters tall. They also know that Alpha, one of the two goddesses, lives on the coldest mountain. And they know that the goddess living on the tallest mountain throws down manna.

We can now argue that this kind of knowledge doesn't allow them to interact with the things they have knowledge of. For instance, Alpha can't know how she would go about climbing Olympus Mons even if she had additional facts in her epistemic pocket—for instance knowledge of the fact that Olympus Mons is on Mars at approximately 18.65°N 226.2°E, that Mars is at least 54.6 million kilometres away from Earth, and so on. Why not? Because she wouldn't be capable of locating Olympus Mons relative to herself. She might have knowledge of the relation between Olympus Mons and other things in the world, including Alpha, but she neither knows that she's Alpha nor where she herself is.

As we saw, Alpha's omniscience doesn't come with a neat little tag labelled 'You are here!' She lacks *de se* knowledge of her own location in the world she thinks about. Hence, her omniscience and perfect

impersonal way of thinking about space doesn't enable her to locate herself in her world. Her knowledge doesn't include knowledge of her own location in the required *de se* sense. But if she doesn't know where she herself is within the context of her thinking about space in this impersonal way, she can't use this kind of thinking in order to interact with the things that are in that world. While she might know the exact coordinates of Olympus Mons, she unfortunately doesn't know her own coordinates. Hence, an allocentric, impersonal way of thinking isn't sufficient to guide interaction with the world.

The same applies if we allowed the goddesses's omniscience to range over certain perspectival facts of the world. For instance, their knowledge could include perspectival relations between things in the world such as one object being to the left or right of another from certain points of view. So, they could know that K2 is to the right and behind Gasherbrum II when viewed from the top of Gasherbrum I. Since perspectival facts are perfectly objective, the gods' omniscience should incorporate knowledge of them. But, while this kind of knowledge enables our goddess Alpha to know how to get from Gasherbrum I to K2, it doesn't help her to get to K2 from where she's at right now. That's because she simply doesn't know who or where she is. For all she knows, she might already be on K2. Hence, thinking about space either in the allocentric, impersonal way, which doesn't include a privileged point of view, or in the perspectival way, which involves points of view but no reference to one's own location, is insufficient for the subject to know how to interact with the things thereby thought about.

Let me formulate this in more general terms as a kind of argument. The impersonal way of thinking about the world that's constitutive of the omniscience of our goddesses entails an equal footing of every location in that world. Hence, such a way of thinking is constitutively without a privileged point of view or centre. The world of allocentric thinking doesn't revolve around one point in space. However, for a subject to be capable of interacting with the world she thinks about, she has to know about her relations to the things she wants to interact with. This is because she needs to know how to move in order to act on objects in the world. Such kind of knowledge requires the subject to locate herself in that world. But, to locate oneself in the world one thinks about requires the materialisation of a privileged point—i.e. the location of oneself, the acting subject. The acting subject's world has

to be centred. Hence, the impersonal allocentric way of thinking of our goddesses isn't sufficient to enable interaction with the world.

Moreover, even if the subject is equipped with some perspectival knowledge about the world—such as knowledge of the fact that K2 is *to the right* and *behind* Gasherbrum II when viewed from the top of Gasherbrum I—such knowledge doesn't entail that the subject knows where she herself is located. Despite the fact that perspectival thinking involves some privileged point, not any point will do for successful coordination between perception and acting. The subject has to take herself to be located at the *centre* of perspectival space and not just anywhere within that space. But one can think perspectivally about space without taking oneself to act from the centre of this perspectival space. Hence, such a perspectival way of thinking about the world enables interaction with the things thought about only conditional on a further premise which is egocentric in nature—e.g. that the subject takes herself to be on Gasherbrum I, the centre of that perspectival space.

These arguments support our first claim that an egocentric way of thinking about the world is necessary for a subject to know how to interact with the things she encounters in the world she thereby thinks about. The merely perspectival way of thinking, involving such terms as *left*, *right*, *above*, *behind*, doesn't provide us with the desired two-way interaction that we witnessed in the case of the painter. What we need is egocentric thinking as the glue that holds together thought about the world and interaction with it. Of course, this idea wasn't just conjured out of thin air. We can trace it back to at least Evans's *The Varieties of Reference*, who puts it very succinctly:

> Egocentric spatial terms are the terms in which the content of our spatial experiences would be formulated, and those in which our immediate behavioural plans would be expressed. This duality is no coincidence: an egocentric space can exist only for an animal in which a complex network of connections exists between perceptual input and behavioural output. A perceptual input (…) cannot have a spatial significance for an organism except in so far as it has a place in such a complex network of input–output connections.
>
> Evans 1982: 154

This quote further consolidates the necessity of egocentric thinking for action. The way we think about the world can only be of significance if it's put in relation with our possibilities of interaction. Furthermore, the quote hints at an explanation of why we should accept the second claim: The egocentric way of thinking is constituted by the subject's assumed possibilities of interaction with the objects in her world. Evans's argument is along the same lines by establishing a constitutive link between how a subject experiences the world around her and how she would interact with the things she thereby thinks about.

If we accept this connection, a particular object we think about egocentrically only gets its characteristic properties on the basis of the two-way relation between thinking about and interacting with the world. That means that egocentric thinking is characteristically tied to our possible actions. Accordingly, our subject thinks of K2 as being to the right and behind Gasherbrum II because she has to move in a specific way in order to reach K2 from wherever she is right now. For instance, she has to first pass Gasherbrum II before she gets closer to her goal of climbing K2. Similarly, an object is thought of as being *within reach* because the subject takes herself to be able to grasp it without, say, moving to a different location. This latter assumption about the possibility of interaction with the object thus constitutes how our subject thinks about the world around her.

Given that egocentric thinking forges the link between a subject's experience of the world around her and her interaction with the objects she thinks about, we may ask: How is it enriched through this active aspect? The answer lies in the necessary requirement of some reference to the subject's assumed possibilities of interaction with the objects. A mere perspectival image from a certain point of view isn't sufficient because it doesn't present the world in a directly engaging way. The presentation of something as being *to the left* does not entail that the subject takes herself to be capable of interacting with that thing by moving in a particular way. This is why it's perfectly perspicuous to perspectivally think about K2 as being to the right of Gasherbrum II without thereby taking a stance on where these mountains are in relation to oneself—and *ipso facto* without thereby taking a stance on how one could interact with these objects. In contrast, egocentric thinking involves the subject's grasp of how she can directly interact with the things in her world.

We may accept the following example as initial support for this thesis. When I am playing tennis, I think about the ball as coming towards me. This kind of thinking is egocentric only if I take myself to be capable of hitting the ball on the basis of thinking about it in that way. My assumed possibilities of interaction—e.g. my assumption that by moving my right arm in a certain way I can drive the ball down the line—are constitutive of how I think of the tennis ball. Without them, I wouldn't think about the tennis ball in an egocentric way. Without them, I would merely think of the ball in the detached way that's characteristic of allocentric or purely perspectival thinking.

Naturally, there are borderline cases which depend on whether the subject really assumes some possibility of interaction with the objects surrounding her or not. For instance, if I watch a video of Roger Federer playing tennis from his own perspective, I might also think about the ball as coming towards me. After all, it looks as if it's coming towards *me*. However, I usually don't take myself to be capable of interacting with the ball on the basis of thinking about the ball in that way. Why not? To give but two examples, I don't expect to be hit if I don't move out of the way of the oncoming ball and I don't expect to be able to drive the ball down the line if I move my right arm in a certain way. These things are by no means necessary. A subject can take up Roger Federer's perspective in an egocentric way and assume that she can hit a winner on the basis of her thinking about the tennis ball—and this assumption would then be deeply disappointed because there's no ball to be hit. This example shows how close egocentric and perspectival thinking are related. One and the same visual image can be egocentric in one case and merely perspectival in the other. What distinguishes the cases is only whether the subject assumes some possibility of interaction (A.4.4).

There's a further dialectical reason why we should accept the claim that egocentric thinking is constituted by a subject's assumed possibilities of interaction—and you might disagree on the conclusiveness of that reason. Remember that we want egocentric thinking to be a key to understand *de se* thinking. As such, it better help us explain some of the characteristic features of our ability to think in the *de se* way. One of these features is that *de se* thinking is necessary to move subjects to action. It's because a subject takes a fact in the world as a reason *for her* to do something that she's motivated to act on the basis of that reason.

I argued that a mere impersonal reason in isolation—such as the mere fact that Olympus Mons is 21'230 metres tall—doesn't move a subject to action because it doesn't constitute a reason for her to do something. Only by taking a *de se* stance on these facts or by conjoining them with additional *de se* mental states will these facts promote behaviour.

If we take egocentric thinking as being constituted by a subject's assumed possibilities of interaction, we can nicely accommodate this important feature. The reason is quite simple. Because a subject thinks of objects in terms of what *she herself* can do with them, she takes what she thinks about as potential reasons for her to interact in some way with these objects. The painter thinking of the canvas as being in front of her provides her with a potential reason to paint by moving her arm in a certain way and not in another. Hence, there's a neat explanatory connection between the proposed concept of egocentric thinking and a key feature of *de se* thinking. If egocentric thinking is constituted by the subject's assumed possibilities of interaction with the objects in her world, she's immediately capable of taking facts in the world as reasons for herself to act in some way or other.

We're now equipped with good reasons to accept the two crucial claims about egocentric thinking that called for support. I argued that egocentric thinking is the right way to account for how subjects find themselves in the world because it's the only kind of thinking that ensures the necessary two-way relation between thinking about the world and interacting with it. Furthermore, we saw that egocentric thinking has to be characterised through the subject's assumed possibilities of interaction with objects presented to her. This is partly because taking facts in the world as reasons for yourself to act in a certain way is conceptually and explanatorily closely connected to thinking of the things in the world in terms of what you take yourself to be able to do with them—an essentially *de se* way of thinking.

And after having explained and justified these claims, we're now in a position to answer the various questions about egocentric space that we started with. The first question was about what constitutes egocentric thinking. The answer is that egocentric thinking is constituted by the subject's assumed possibilities of interaction with the objects thus thought about. These assumptions needn't come in the form of explicit judgements such as *This tennis ball can be hit by me in such and such a way*. Rather, they're manifest in what a subject takes herself to

be capable to do on the basis of thinking about the world in that particular way. Furthermore, we saw that egocentric thinking comes with a clearly prioritised location from which the subject interacts with the world around her. This is the *origin* of egocentric space: the conceptual and behavioural centre of a subject's world.

Our second question concerned the relation to other ways of thinking about the world. I argued that these constitutive features distinguish egocentric thinking from other ways of thinking about the world. Neither perspectival nor allocentric thinking is characterised in terms of the two-way relation between thinking about and interacting with the world *and* includes a prioritised centre from which the subject interacts. This distinction and clarification provided us with the argumentative fuel to defend the thesis that subjects are capable of directly interacting with the objects in the world only by thinking about space egocentrically. And we thereby answered the final question: Why do we have to characterise the way a subject thinks about her world as egocentric? If objects are thought about with direct reference to what the subject takes herself to be capable of doing with them, these subjects immediately know how to go about interacting with them (A.4.5).

Now, how is this related to *de se* thinking? Most of the examples and arguments involved cases of thinking about the things out there in the world and not about the thinking subject herself. But, of course, a subject thinking egocentrically about the world can also thereby think about herself. One of the discussed examples was the painter who thinks about her arm and believes *My arm is bent* on the basis of feeling the bend in her arm. In such a case, we can say that the arm is given to her in a different way than the brush or the canvas. Rather than being an independent object with which she can interact, her arm belongs to the origin of her world *through* which she interacts. This origin takes up a unique epistemic role in her egocentric thinking. It constitutes the thing through which she's able to interact with the objects in the world. The arm thus belongs to that primitive basis of how things are given to her in egocentric thinking—as *graspable*, *within reach*, or *an arm-length away*. But what exactly is the origin of egocentric space? What's this 'thing' with which the subject feels, grasps, sees, desires, believes (A.4.6)?

4.2 THE LIVED BODY

The fascinating case of the rubber hand illusion provides a graphic starting point to explore the nature of the origin of egocentric space. It's a rather easily produced illusion where the subject takes an artificial hand to be her own real hand. For instance, she feels touch and movement 'in' the rubber hand, flinches if a hammer is brought down on the rubber hand and points towards the rubber hand if prompted to touch her own hand. The illusion was first produced and reported by the psychologists Botvinick and Cohen (1998) and consists of a simple setup. A subject's left hand is placed on a small table and hidden from direct view by a screen. Then, a rubber hand is placed in front of the subject and she's instructed to fix her eyes on the artificial hand. Finally, both the real hidden hand and the artificial visible hand are stroked synchronously by two brushes for up to ten minutes.

What happens now is that the subject starts to associate the brushstrokes she *feels* on her real hand with the ones she *sees* on the artificial hand. This association produces the illusion of 'feeling' the touch of the brushstrokes in the rubber hand. Furthermore, if the subject is asked to point blindly to her left hand with the index finger of her right one, she will point closer to the artificial hand than to her real one. The result is an illusion in which the subject experiences her left hand as being located where the artificial hand is.

Things get even more interesting in a slight variation of the illusion which is a more recent development (cf. Kalckert and Ehrsson 2012). Here, the rubber hand is placed directly above the left hand on a small platform. Then, both index fingers—the hidden real one and the visible artificial one—are connected with a rigid stick. After stroking both hands for a while, the subject reports to feel the brushstrokes in the artificial hand just as in the original setup. But while the original illusion is easily undermined and broken by moving the finger of the hidden hand and not seeing the expected result in the rubber hand, the alternate setup allows for an even deeper immersion into the illusion. If the subject now moves her left index finger up and down, she sees the rubber index finger moving accordingly and thus has the illusion of moving the rubber hand through her will. In other words, not only does the subject 'feel' touch in the rubber hand, she might also assume that she can interact with the world through the artificial hand instead

of her own real hand. This produces an even more immersing illusion of the rubber hand being one's own.

How do we explain what's going on in this interesting illusion? One scientific possibility is to call upon the integration of information from different sense modalities that occurs when subjects interact with the world. We normally have to put together massive amounts of information which we receive from ourselves and the objects surrounding us into a coherent picture. For instance, we combine information we receive from smell, touch, and vision to form a coherent perception of a rose under our nose. Now, in the case of our illusion, the subject has to make sense of the different inputs she receives from vision, touch, and proprioception. By carefully capitalising on the way we usually integrate this input, we can produce the rubber hand illusion. The real hand is blocked from sight while the feeling of touch and proprioception is instead synchronised and thereby linked to the vision of the artificial hand. This produces a deviant integration which fools the subject into taking the rubber hand as her own. The synchronicity of the brushstrokes and the movement of the fingers is thus capable of deceiving us into wrongly unifying inputs which originate in different parts of egocentric space.

There's a problem with this explanation though. While it might be scientifically well-established and accurately accounts for what's going on while a subject undergoes the illusion, it doesn't exactly explain what it means for the subject to take the rubber hand *to be her own*. While it explains why and how the cognitive system builds the 'hypothesis' which can be expressed in the subject's belief *This is my hand*, we don't know what it means for the subject to have that particular belief. The genesis of such a belief in the cognitive system is one thing, the meaning and significance for the subject quite another. While we know how that belief was generated in that particular instance, we don't know how a subject takes *anything* to belong to herself. So, maybe a less sophisticated starting point is better. We might just want to say that subjects experience 'the rubber hand as belonging to themselves' (Botvinick and Cohen 1998: 756).

Of course, this isn't an explanation either. We have to give an account of what it means for a subject to take something as belonging to herself. After our discussion of the origin of egocentric space, the following proposal shouldn't come as a surprise: To take something as

belonging to ourselves is just to take something as belonging to the origin of egocentric space. What does this amount to? Human subjects normally take their hands to belong to the origin of egocentric space. They assume that they're able to grab things with their hands, point to objects with their fingers, catch objects thrown at them, and so on. However, in the case of the illusion, the subject could be said to erroneously take the rubber hand to be the thing with which she's capable of doing all these things. She assumes that she can interact with the world around her through the artificial hand while erroneously taking her real hand *not* to belong to the origin of egocentric space. And this just is what it means to take the rubber hand to belong to oneself. Because the subject accepts it as belonging to the origin of egocentric space, it's accepted as belonging to herself.

We see, then, that the rubber hand illusion can be taken as a prime tool to illuminate the specific concept of origin which is in play in the concepts of primitive self-ascription and egocentric space. In the illusion, the subject assumes that the rubber hand constitutes a possibility of interacting with the world. For instance, if there was a fly sitting on the artificial index finger, she would move her index finger up and down to shy it away. This corresponds to how she would interact with the world outside of the illusion. Moreover, if she would want to grab a glass in front of her, she would move her hand *as if* it were located where the artificial hand is. The rubber hand thus can be said to belong to the origin of egocentric space: the epistemic and behavioural centre of a subject's world. In this context, it's important to distinguish between actual possibilities of interaction and the subject's assumed possibilities of interaction. A subject can assume to be able to grasp the glass with the artificial hand even if that's not actually possible. How so? She might move her actual hand in such a way that, were it located where the rubber hand actually is, it would end up where the glass is. As such, she assumed that she can interact with the glass through the artificial hand because her movement had its intentional origin in the location of the rubber hand. If she had assumed otherwise, she would have moved her actual hand differently. This tells us that her assumptions shouldn't be understood as fully conceptualised judgements that she actively forms and sustains. Rather, we have to understand them as being part of a network of possible and actual inferences as well as possible and actual actions.

How is this connected to the subject's ability to self-ascribe properties and entertain *de se* beliefs? Let's take the following specific belief which the subject forms while undergoing the illusion:

(13) *My finger is bent.*

The crucial question is: How does the subject come to self-ascribe the property of having a bent finger in believing (13)? The reason for her self-ascription might be partly found in a specific proprioceptive feeling of a bent finger. Additionally, that experience is coupled and integrated with her seeing the artificial finger as being bent. On that complex epistemic basis, she takes the rubber hand to be a source of her behaviour in the world, a thing through which she can directly interact with the world. And this is exactly the role which is played by the origin of egocentric space.

Thus, the origin plays an anchoring role for primitive self-ascription and *de se* thinking. Her taking the rubber hand to belong to the origin of her world produces the capability of primitively self-ascribing the property in question. There's no need for her to identify her finger *as* her finger within the experienced world. Quite the opposite: the finger belongs to the origin of egocentric space. And since there can't be more than one such origin, identification isn't required. Rather, the finger is given to her in an epistemically primitive and direct way as the thing through which she acts. In other words, what it means for the subject to take the finger as belonging to herself isn't the identification of that finger *as* her own but rather the finger constitutively belonging to the subject's origin—through the subject behaving as if the artificial hand were a possibility of direct interaction with the world. A related point is that the rubber hand illusion tells us something about the sense of ownership. The reason the hand is experienced as one's *own* isn't that the subject judges it to be *hers*. Rather, it's the fact that it belongs to the origin of her world which underlies any such possible *de se* judgement.

Putting this all together, we can conclude that the origin provides the epistemic basis on which the subject comes to believe (13) by primitively self-ascribing the property of having a bent finger. And this is nothing over and above the ascription of that property to the origin of egocentric space. This again is enabled by the subject taking the finger to belong to that origin. Primitive self-ascription, which forms the basis of *de se* thinking, is thus grounded in that particular concept

of the origin of egocentric space. A subject's experience of the world from the origin of her world brings about the capability to primitively self-ascribe properties and thus creates the possibility of *de se* thinking in that it explains how subjects come to grasp that they're ascribing a property to *themselves* in the first place.

The case of the rubber hand illusion illuminates further that the concept of the origin is connected to a subject's assumed possibilities of interaction with the world. It's because the subject is fooled into thinking that she's able to shy the fly away with the artificial hand that we can understand her as taking that hand to belong to the origin. We might have initially thought that the physical body plays the role of the origin of egocentric space. But this proves to be the wrong approach. In the deviant case of the rubber hand illusion, the subject takes the artificial hand to belong to the origin. She assumes that she can grab the glass with that hand, she assumes that she feels a gentle touch on its skin and so on. Thus, the origin doesn't just correspond to the physical body. Rather, following some philosophers in the phenomenologist tradition, we should call it the *lived body*.

What is the lived body and what role does it play in our story? As a first approximation, we might say that it's the malleable, living, and experiencing body through which a subject interacts with the world. The lived body is phenomenological in nature and intended to play a foundational role for *de se* thinking. To be more precise, it explains how primitive self-ascriptions are possible. This is because the account under development claims that primitive self-ascriptions are ascriptions of a particular unique kind: they are ascriptions to the lived body—the origin of egocentric space. If we understand primitive self-ascription in this way, it implies that there's no reliance on any prior epistemic achievement such as the identification of a particular body as one's own. Rather, something being taken as one's own body is nothing over and above something belonging to the lived body.

Of course, this is too broad to serve as a good account. So, how can we further specify the concept of the lived body so that it can play this particular foundational role? The concept of the lived body is certainly nothing new, so we can look to established theories for help. In the short synopsis of my account in chapter 1, we already witnessed that Edmund Husserl, the founder of phenomenology, provided the basis for the concept of the lived body. But it finds its maybe most

pronounced discussion in Maurice Merleau-Ponty's *Phenomenology of Perception* (2012); for instance in the following passage:

> I only foresee what this form [of stimulation] might be by leaving behind the body as an object, *partes extra partes*, and by turning back to the body I currently experience, for example, to the way my hand moves around the object that it touches by anticipating the stimuli and by itself sketching out the form that I am about to perceive. I can only understand the function of the living body by accomplishing it and to the extent that I am a body that rises up toward the world.
>
> Merleau-Ponty 2012: 77–78

The situation under discussion in this passage is the specific kind of experience a subject undergoes when her arm is touched. Merleau-Ponty argues that it's one thing to describe the situation from an 'objective' outside perspective and quite another to specify the way the subject thinks about her arm in an egocentric experiential way. He argues that we can only understand the general way in which touch is experienced by a subject by moving away from a third-personal scientific explanation to a description in terms of the subject's experienced lived body. In other words, touch is a kind of experience that's essentially tied to a subject taking a particular first-personal stance on the world through her lived body—a body which isn't established as an object among others but as a privileged living and experiencing thing.

But how can we establish such an argument? We have to focus on and justify the idea that the experience of touch isn't exhausted by merely listing the physical processes which are involved when, say, a feather is brushed against a human body. While it's certainly possible to explain *something* by providing such a story, we're missing an important aspect without which that kind of situation isn't an *experience* of touch at all. We could say that the experience of touch is more than the mere physical contact of a feather against human living skin.

But what is this missing aspect? It's the seemingly simple fact that the subject takes the touch to be experienced in *her* own body. The merely scientific story—which treats the body as an object among other objects—doesn't provide this because it doesn't clearly distinguish between the cases where a subject experiences something as hap-

pening to herself and a subject that undergoes a certain experience without attributing that experience to herself. Touch is the experience of feeling something brushed against *one's own* body. And an 'objective', physiological story is bound to overlook this qualitative, subjective aspect. This also shows that the phenomenon that Merleau-Ponty wants to isolate isn't the mere *consciousness* of the experience of touch. Rather, it's the experience of *being touched* that's our point of interest. We might consciously experience something in our physical body without necessarily taking the experience as occurring in our lived body and thus as being an experience of ours.

This argument carves out the first important feature of the concept of the lived body. While the physical body of a subject can be fully characterised by giving a third-personal scientific explanation, the same isn't true for the lived body. The two kinds of bodies might often coincide spatially, but the lived body is constituted through the first-personal experience of assumed possibilities of interaction with the world around us. In that sense, the lived body isn't on the same conceptual and phenomenological level as the physical body. The latter is primarily given to the subject as one object among many others in the world, while the former is given as the privileged body through which the subject takes herself to be capable of moving around her world, experiencing it and interacting with the objects around her. It's in this sense that the lived body metaphorically 'rises up toward the world'—the subject *engages* with the world through the lived body—whereas the physical body is a mere object among other objects.

Our discussed case of the rubber hand illusion clearly illustrates this difference between the physical and the lived body. On the basis of taking the artificial hand to belong to the lived body—independently of taking it to be part of the physical body—the subject is capable of self-ascribing a certain property. She assumes that she can grab things with that hand, catch things, experience touch, and so on. And these assumptions aren't tantamount to particular explicit judgements but manifest themselves in how the subject engages with her world. As such, the lived body is the thing which is epistemically given to her in a direct and primitive way. And this is why she needn't pick out and identify her hand *as* her hand in the objective world. But certainly, the rubber hand isn't part of her physical body. So, her lived body can't be the same as her physical body. At the same time, the actual hand,

despite being a part of her physical body, shouldn't be described as belonging to the lived body while she's affected by the illusion. It's located 'outside' of her assumed possibilities of interaction. In other words, it isn't taken as a thing through which she can manipulate and experience the world and as such isn't taken by the subject to belong to herself.

We can thus formulate a first distinguishing feature of the lived body, which corresponds to the subject's origin of egocentric space. The lived body is distinct from the physical body. Thus, a subject might take something as belonging to the lived body without it constituting a part of her physical body and *vice versa*. We have to understand the lived body as a phenomenological notion which is grounded in a subject's engagement with the world around her. On the other hand, the physical body is a purely 'objective', scientific notion which is wholly established through underlying physical processes (A.4.7).

Two further constitutive features of the lived body were already implicitly mentioned. One of them is the fact that a subject's lived body is established through what she takes to be possible ways of interaction with the world. This follows from the fact that the lived body is an elucidation of the concept of an origin. I argued that the origin of egocentric space is the node which ties a subject to her world. As such, the assumed possibilities of interaction play a crucial role because they establish the prospect of engagement with the world. Accordingly, if something is taken by the subject as a direct means of interaction with objects around her, it belongs to the origin of egocentric space. And because the lived body just *is* the origin of our world, the same applies to it. Whatever is taken by the subject—through her way of interacting with the world—as a way of direct interaction with objects in egocentric space belongs to the lived body. For instance, my fingers belong to the lived body through which I engage with the world. They enable me to interact with the keyboard in order to type these sentences. They are the medium through which I feel the pressure of the keys and the warmth of the sun shining on them. I assume, through my engagement with the world, that these fingers are a possible way for me to interact with the world. And this is why they belong to the origin of egocentric space and thus to the lived body.

The case of the rubber hand illusion is a deviant though analogous example. While under the illusion, the subject might take herself to

be able to grab things with the artificial hand but not with the real one. The real hand might be the actual thing with which engagement with the world is possible, but the subject assumes differently. Similarly, an artificial limb might be experienced as belonging to the lived body because it constitutes part of the node between the subject and the world: It enables the subject to walk around, kick things, stumble, and so on. The lived body is thus formed through the subject's assumption about how she's capable of directly interacting with the world. Keep in mind, though, that this assumption isn't a conscious and active stance towards the lived body and the world. Rather, it shows itself in the engagement of a subject with her world. Hence, my assumption that I'm capable of interacting with the world through my fingers shows itself in my engagement with the keyboard and not necessarily on the basis of my judging *I can interact with the world through these fingers*.

Merleau-Ponty provides us with a nice example which further illustrates this important and constitutive feature of the lived body. By looking at the curious phenomenon of the phantom limb, he argues that we have to understand this case as a subject who 'refuses' to accept the loss of some of her possibilities of interaction with the world. Despite visually seeing and accepting the absence of her arm, she takes it to belong to the lived body, the body through which she engages and interacts with the world. Once more, this illustrates the distinction and potential divergence of actual and assumed possibilities of interaction. Moreover, the phenomenon emphasises the underlying and latent nature of these assumptions. Maybe the most conspicuous expression of this analysis can be found in the following passage:

> To have a phantom limb is to remain open to all of the actions of which the arm alone is capable and to stay within the practical field that one had prior to the mutilation. The body is the vehicle of being in the world and, for a living being, having a body means being united with a definite milieu, merging with certain projects, and being perpetually engaged therein.
>
> Merleau-Ponty 2012: 84

Of course, this important feature of the lived body doesn't come as a surprise and is nothing new. We've already established that the origin of egocentric space is defined through the subject's assumed

possibilities of interaction with the world. And since the lived body is supposed to function as that which illuminates and constitutes a subject's origin, it has to be defined in the same way. Because of that, let me just briefly repeat the previous arguments for this feature without further commentary. We witnessed that the requirement for the lived body to establish a subject's active engagement with the world makes it necessary to incorporate the two-way relation of egocentric thinking. And the claim that the lived body is established through a subject's assumed possibilities of interaction is a reformulation of this point. Furthermore, this understanding of the lived body puts it conceptually and explanatorily close to a crucial feature of *de se* thinking: our ability to think of the things in the world in terms of what we take ourselves to be able to do with these things.

The third feature of the lived body equally follows from our discussion of the nature of the origin of egocentric space. We should understand the lived body as essentially first-personal. Let me explain the reason for and meaning of this feature. The general problem of *de se* thinking is how we account for the capability of subjects to think about themselves in such an intimate way, allowing them to be aware of the fact that they're thinking about *themselves*. In other words, we want to explain the essential first-personal nature of *de se* thinking. The arguments for the requirement of primitive self-ascription showed us that we have to accept a certain basic form of *de se* thinking—i.e primitive self-ascription as ascription to the lived body. Every other way of thinking about oneself in the *de se* way—such as our use of the first person concept or everyday speech in the first person—has to be built on this foundational first-personal capacity.

A short argument for this claim will freshen our memory. Our ability to think about ourselves using the first person concept is a first-personal ability. That's to say that we're aware of the fact that we're thinking about ourselves on the basis of our knowledgeable use of that concept. But why is that ability first-personal? If the arguments so far have been correct, we have to look for something underlying our use of that concept which plays the primitive first-personal role. This is because the referential rule governing our use of the first person concept doesn't account for the first-personal nature of *de se* thinking. A subject can understand that every application of the concept refers to the thinking subject without thereby grasping that her own application of

that concept refers to herself. We then established the required primitive foundational basis to be the capacity of subjects to ascribe properties to the origin of their egocentric space. And this is tantamount to an ascription of a property to the lived body.

But why should we think that the lived body is the wanted essentially first-personal basis? Well, there are two possibilities. Either the lived body is first-personal, or it isn't. Let's imagine it weren't. If that were the case, we would be left with the ascription of properties to the lived body. However, nothing about this would be first-personal. Neither the lived body would be first-personal nor the ascription to it. Why not? Because our account explains the essential *de se* nature of primitive self-ascription by claiming that it's an ascription to a very special thing: the lived body. What makes a self-ascription primitive is the fact that it's a self-ascription to the lived body. But if the lived body isn't first-personal, then neither is primitive self-ascription. Then whence does the essential first-personality of *de se* thinking come from? By characterising the lived body as essentially and basically first-personal, we have a natural basis of that feature. In other words, the lived body's first-personal nature explains why ascription of properties to it amount to primitive self-ascriptions, which form the basis of *de se* thinking (A.4.8).

Finally, the fourth and last feature is closely related to the previous one and can be derived from the same considerations. The lived body has to be given to the subject in a non-representational way. That's to say that the subject can't think of the lived body in a certain way *as* something or other. To better understand what that means, compare the way a subject thinks of the lived body to the way Alpha thinks of the tree in front of her *as* an oak. In that case, Alpha determines that what she perceives has certain features and qualities and thus judges it to be an oak. The claim is now that we can't and shouldn't tell the same story about the lived body.

Let's look at how we can justify this claim. There are two main reasons against the idea that a subject has to think of the lived body in a certain way—for instance as *her own*. The first one is that our account shouldn't leave room for the possibility of wondering whether the lived body that's given to the subject is her own or someone else's. Remember that the lived body is the exact means to dissolve that doubt. It has to enable a subject to interact with the world without any prior

epistemic achievement. And thinking about the body *as* 'one's own', 'mine', or 'the body I inhabit' exactly amounts to such an undesired epistemic achievement.

Let me be clear. There's a sense in which we have to express a subject's engagement with the world through her lived body as her taking the lived body 'as her own'. However, this is only a superficial relation between the subject on the one side and the lived body on the other. There's neither a real metaphysical nor epistemic relation corresponding to it. It's merely a perspicuous and necessary, but deceptive, way of speaking and writing. The subject can't literally think of the lived body as her own. Rather, we explain what it means for a subject to accept a given body as her own by saying that it constitutes her lived body. This doesn't mean that any belief of the kind *This body is mine* is deceptive or false. Quite the opposite. We have to understand her grasp of taking a body as hers as being based on her lived body in a non-representational way. Furthermore, the body she's thereby thinking about isn't the lived body but the physical one.

The second reason is that a subject would have to identify a particular body *as* her own if she had to think of the lived body in a certain way. But this introduces two problems: It opens up the possibility of misidentification and it leads into a potential infinite regress. The lived body is the rock bottom foundation for *de se* thinking and as such has to provide the means to account for the characteristic features of thinking in the *de se* way. One of these features is the potential immunity to error through misidentification. But if the lived body isn't equipped with the required basis for immunity, then it's difficult to see how *de se* thinking could be exempt from misidentification. Furthermore, if we assumed that the subject had to identify the lived body as her own, we might ask: Doesn't that require a further identification? For, what does it mean to identify something *as one's own*? We couldn't muster the lived body as an explanation anymore, hence something else would be required. But what keeps us from asking the same question one level down? As long as we think of the foundation of *de se* thinking in identificational terms, this question arises again and again. Hence, these considerations speak for the fact that the lived body is free from identification.

This concludes the elaboration of the proposed positive account of primitive self-ascription. We're now left with a good picture of the nature of the origin as the lived body. In order to provide a founda-

tion for *de se* thinking, we ventured into the concept of primitive self-ascription. I then argued that primitive self-ascription has to be understood as ascription to the origin of egocentric space. This again called for a picture of what that mystical origin is. The lived body provides a good basis which, if understood properly, is capable of playing the foundational role for *de se* thinking. Let me summarise this in the form of a definitive characterisation:

THE LIVED BODY ACCOUNT OF DE SE THINKING
When a subject thinks about herself in the *de se* way, she entertains a thought that's partly constituted by her ascribing a property to the lived body.

This characterisation is only meaningful if complemented with an account of the lived body. In the course of our discussion, we've identified four essential features. The lived body is (a) phenomenally and conceptually distinct from the physical body, (b) constituted by the subject's assumed possibilities of interaction with the world, (c) essentially first-personal, and (d) free from identification. However, not all *de se* mental states can be analysed in terms of a specific primitive self-ascription with which they correspond and hence, *de se* thinking is only *partly* constituted by ascriptions to the lived body. As we'll see, some of our *de se* thoughts don't directly concern the lived body and we might be able to think in the *de se* way despite a temporary 'absence' of the lived body. However, a thought's *de se* nature has to be traced back to some ascription to the lived body which makes it *de se* in the first place. Now, we're going to put all the puzzle pieces together (A.4.9).

4.3 PRIMITIVENESS AS ORIGINAL THOUGHT

One main objective that remains now is to spell out how the presented account explains various cases of *de se* thinking. Some of the discussed examples already indicated the connection between primitive self-ascription, egocentric thinking, and the lived body as the origin. But now it's time to make these conceptual connections explicit. We may think back to the opening question of this book, which was to examine the nature of our thoughts about ourselves. In the paradigmatic case of Narcissus, we have a highly self-conscious individual who's capable of self-consciously thinking about itself. At some point he be-

comes aware of the fact that he's looking at *himself*. Now, it's his ability to think in the *de se* way which underlies that revelation. And similarly, John Perry's realisation that he *himself* is making a mess in the supermarket capitalises on the general ability to think *de se* thoughts—thoughts which are necessarily about the thinking subject herself.

On our journey to get a grip on this most basic and intimate way of thinking about oneself, we stumbled upon Lewis's and Chisholm's property theory. There, we became acquainted with the claim that *de se* thinking is best understood as being grounded in the ability to self-ascribe properties. Through our discussion of this claim, we specified it further and argued that this has to be understood in an epistemically primitive way. Some of our self-ascriptions have to be basic without depending on any further epistemic achievement. Most importantly, a subject's awareness of the fact that she's thinking about *herself* can't depend on a prior identification of a particular object as herself. These considerations also formed part of the reason to reject other strategies that were discussed earlier.

Finally, I argued that only a phenomenological notion such as the lived body can ultimately illuminate the nature of primitive self-ascription. The intimate tie between the lived body and a subject's engagement with her world makes it the ideal foundation for *de se* thinking. According to the lived body account, a subject can think about herself in the *de se* way because the lived body is given to her in the epistemically basic way required for primitive self-ascription. This entails a certain metaphysical claim which comes down to this: every subject is constituted through her lived body—the origin of egocentric space. It's the nexus of her engagement with the world. And this privileged position allows a subject to ascribe properties to herself in the epistemically direct way that grounds our ability to think *de se* thoughts.

From all this, the following picture emerges: Whenever a subject thinks about herself in the *de se* way, that way of thinking is at least partly based on her primitively self-ascribing some property or other. Primitive self-ascription, in turn, is nothing but the ascription of a property to the origin of egocentric space. In addition, the fact that we're the origins of our worlds enables us to think about ourselves in that epistemically direct way. We occupy a privileged position in our world, which brings about the possibility to think in an intimate way about this most exceptional thing: ourselves.

A good way to illuminate and ponder the merit of such an account is to confront it with some of the paradigmatic kinds of examples that have been discussed in the book so far. By doing this, we will see how the idea that *de se* thinking is partly based on primitively self-ascribing a property accounts for these examples. Furthermore, this will deepen our understanding of both the role of the lived body in *de se* thinking and the nature of primitive self-ascription and its relation to other self-ascriptions. Maybe the most vivid and instructional contrast is between the following two seemingly identical beliefs:

(14) *My back is horizontal* formed on a proprioceptive basis.
(15) *My back is horizontal* formed on a visual basis.

Here we have a contrast between two beliefs of the same kind which are formed on very different epistemic bases. The belief in question is that of a subject taking her back to be in a horizontal position. We can assume an important similarity between the two: the believing subject engages with the world through the lived body in either case. More precisely, this engagement includes some underlying assumptions about how her back relates to the world around her. For instance, she assumes that she can lie on her back in a flat position, that her back is strong enough to lift the box, that she's ticklish in the lower half of her back, that it hurts under her shoulder blades, and so on. In other words, her back belongs to her lived body, which is constituted by her assumed possibilities of interaction with the world around her. However, there are significant differences too.

In the case of Alpha believing (14), she's in a situation where she experiences her back as being horizontal. Normally, human subjects are pretty bad at determining when exactly their back is perfectly horizontal. But we can imagine Alpha as being rather good at yoga, and thus having developed a good feeling for her own posture. Hence, she's capable of proprioceptively judging her back to be parallel to the ground even in complicated bodily positions. In such a case, we'd want to say that the basis of her belief lies in her specific experience of the world through the lived body. Importantly, the possibilities of interaction with the world depend on the specific position the lived body currently occupies relative to other objects around Alpha. For instance, she assumes that, were one to put a snooker ball on her back, it wouldn't roll away. In other words, she assumes that her back is

capable of supporting a rolling object. And this is tantamount to her believing her back to be horizontal.

The experience of the world through her lived body thus allows her to ascribe certain properties to the origin of her world. Certainly, not all properties can be thusly self-ascribed. While she might be in a position to primitively self-ascribe the property of *having a horizontal back*, she most likely isn't in a position to primitively self-ascribe the property of *having a tattoo with the shape of a Pegasus on her back*. The way we imagine Alpha, such a tattoo doesn't figure in the lived body and thus can't be taken as a basis for primitive self-ascription. In contrast, her back's posture is reflected in the lived body through the kinds of interactions Alpha assumes are possible with the objects around her. And thus, her *de se* belief (14) is partly grounded in her ascribing the property of having a horizontal back to the origin of her egocentric space by depending on the nature of the lived body.

If we now look at the belief (15) against this background, we'd tell a quite different story. Imagine that Beta believes (15) on the basis of seeing herself in the mirror. She might be engaging in the same yoga exercises as Alpha. But, as a beginner, she lacks the acute sense of her bodily posture and has to visually confirm that her back is in fact horizontal. We can even imagine her as potentially believing her back to be slightly sloping downwards were she to base that belief on her bodily experience alone. However, the lived body only plays a subordinate role in the case of (15) and the visual identification of herself with the mirror image is in the foreground. We could say that there's a lived body looming in the background and enabling primitive self-ascription. But the belief (15) is formed on the basis of the visual experience of *some* back being in a horizontal position together with the further identifying *de se* belief *This is my back*.

The upshot of this exposition is that the resulting *de se* belief (15) is tantamount to the self-ascription of the property of having a horizontal back. However, in the case of (15), that self-ascription isn't primitive. Rather, it epistemically depends on a further *de se* belief which, in turn, might correspond to a primitive self-ascription or not. At some point in that game of tracing back, we'll be left with a belief that's constituted by the ascription of a property to the lived body. In our case, such a belief would most likely be to the effect of Beta considering her back to belong to the lived body. And this grounds her

ability to think about her back in the *de se* way in contexts which aren't primitive.

The contrast between these two beliefs shows us three important things. First, there might be cases in which a subject has two *de se* beliefs of the same kind which correspond to different self-ascriptions of properties. For instance, Alpha's self-ascription is primitive while Beta's self-ascription isn't. We tracked down the reason for that in the difference between the epistemic bases for the two beliefs. In Alpha's case, her basing the self-ascription on the direct experience of her own body allows it to be of the primitive kind. Beta, on the other hand, forms her belief on the basis of further self-ascriptions and beliefs in order to arrive at the *de se* belief (15). Most importantly, some identificational belief has to be present. Hence, her self-ascription of the property of having a horizontal back doesn't amount to an ascription to the lived body and hence isn't primitive in the required sense.

This shows us, secondly, that there are some proper *de se* beliefs which don't correspond to primitive self-ascriptions of a property. Beta doesn't base her belief (15) on the direct experience of her back as being horizontal. Rather, she comes to her belief indirectly via the combination of seeing the back in the mirror as being horizontal and the belief that she herself is the subject in the mirror. This reliance shows us, finally, that all *de se* beliefs have to epistemically involve some ascription of a property to the origin of egocentric space in order to be *de se* at all. A subject's ability to think of something as *herself* is ultimately established in the lived body and thus through primitive self-ascription. Without primitive self-ascription in the background, no belief can be grasped by the subject to be about herself (A.4.10).

It's because of this dependence of all *de se* beliefs on some form of primitive self-ascription that even quite mundane beliefs really are *de se* in nature without corresponding to primitive self-ascriptions. For, most of our *de se* beliefs don't directly correspond to some *primitive* self-ascription. Let's take the following belief as an example:

(16) *I won the lottery.*

There's normally nothing about a subject's lived body that would provide her with a reason to believe (16). But of course, Alpha could be in a situation where she's holding her winning ticket while verifying the correctness of the numbers and then coming to believe that

she's won the lottery. In such a case, Alpha certainly entertains a *de se* belief. After all, her belief is about herself and she's perfectly aware that it's about herself and nobody else. But her belief is dependent on other epistemic groundwork—such as the belief that these are the winning numbers, that the ticket she's holding between her fingers is the winning lottery ticket and that the fingers with which she's feeling the paper are hers. Hence, ultimately, her believing (16) depends on some ascription of a property to the lived body. This epistemic basis is responsible for a subject's potential awareness of the fact that she's thinking about herself when she's thinking in the *de se* way.

The connection between *de se* thinking and the concept of egocentric space exposes us to kinds of beliefs that are more difficult to explain and make sense of. These beliefs aren't properly about oneself and hence don't seem to be *de se* in nature—after all, their intentional object is something else. But they still only seem to make sense to the believing subject when they're put into relation with herself and her own way of thinking about the world. Let's look at one example of this kind:

(17) *There's food in reach to the left*.

On first sight, it seems obvious that such a belief isn't *de se*. After all, a subject who believes (17) isn't thereby thinking about herself at all. The intentional object of her belief isn't herself but the food in question. So, why should we still think that it's *de se*? We can identify two reasons. The first is semantic in nature and the second epistemic. Let's discuss them in turn. Semantically, the conditions of satisfaction of believing (17) depend in the very same way on the believing subject as in more typical *de se* beliefs. If Alpha entertains (17), the belief is true if there's food to the left of *Alpha*. Conversely, if Beta entertains the same belief, her belief is true if there's food to the left of *Beta*. This tells us that while the believing subject isn't clearly visible in our linguistic expression of the belief, it's nonetheless implicitly part of the belief. In order to get the right conditions of satisfaction, we need to make reference to a specific thinking subject. Hence, egocentric beliefs like (17) are only semantically complete if we take into account who's entertaining that particular belief (A.4.11).

We can reinforce this conclusion by looking at the epistemic behaviour of these egocentric beliefs. This will lead to the claim that beliefs

like (17) are in some sense epistemically based on the self-ascription of a property. To put it more precisely, they're dependent on a subject thinking about the world in egocentric terms. But a subject can only engage with the world egocentrically if she thinks about objects in terms of her own possibilities of interaction with them. And this brings the involvement of primitive self-ascription with it. The dependence and involvement can be demonstrated by discussing two terms we find in the belief (17). First, what does it mean for a subject to think about an object in egocentric terms as being 'to the left'? And secondly, what does it mean for a subject to think about an object as being 'in reach'?

Let me start with the latter question. Whether the food is actually in reach for Alpha isn't dependent on her thinking about the world in a particular way. It's a perfectly determinable relation which either obtains between the two things or it doesn't. But, as soon as Alpha thinks of an object as being in reach, she assumes a certain possibility of interaction with that object. For instance, she assumes that she can touch the food if she extends her arm. And this is a case of a subject self-ascribing a property. She's implicitly bringing the lived body, according to which grabbing the food is an assumed possibility of interaction, into the picture. And this forms the basis for her primitive self-ascription of a property. Hence, thinking of the food as being in reach involves some *de se* element by epistemically depending on an ascription to the lived body.

Now, the same applies to the first question. Either the food is to the left of Alpha or it isn't. That much doesn't depend on any self-ascription. But as soon as Alpha thinks of the food as being to the left, she thinks of it as being to the left of *her*, which involves the self-ascription of some property. Thus, even in those beliefs which aren't *de se* on the surface, but involve egocentric terms, there's some underlying self-ascription looming in the background. And therefore, we should say that they're only meaningful to a subject if there's some *de se* state lending its support (A.4.12).

There are two more kinds of *de se* beliefs which I want to discuss. The first kind concerns cases of inserted thoughts, where a subject entertains a belief about herself but doesn't accept that she's the source or agent of that belief. The second concerns those self-ascriptions which are simply false. Let's examine the first class of inserted beliefs by looking at the following example:

(18) *I can fly*.

We should begin with a short message of caution. The psychological phenomenon of thought insertion is extremely complex with conceptual and explanatory problems left and right. As such, this short discussion shouldn't bear much argumentative and deciding weight on neither the proposed account nor on the phenomenon itself. Rather, it's supposed to point into a possible direction in which the proposed account can explore these special mental states. With that said, what does it mean for a thought to be inserted? In the case of Beta experiencing her belief (18) to be inserted, she would claim a lack of agency or authorship of that particular belief. She might explain that the thought isn't her own but instead popped up in her mind or was inserted there by some other powerful subject. But certainly, she would deny that she actually believes that she herself can fly—despite 'believing' (18).

Given this scenario, we might ask whether that belief is *de se* at all. A positive answer is suggested by the fact that the belief has the semantic structure of a *de se* belief. There's no obvious difference between Beta's belief and Wendy's belief that she can fly. Both are true if the believing subject can in fact fly—which is true in Wendy's case but not in Beta's. And why should the belief (18) in Wendy's mind be *de se* but not in Beta's? They're beliefs of the same type. Maybe looking at the specific epistemic features underlying Beta's entertainment of (18) can shed some light on the matter. We can reasonably doubt that (18) exhibits the characteristic features of *de se* thinking in Beta's case. It isn't immune to error through misidentification because she doesn't think that the belief is about her at all. It doesn't serve as a basis for self-knowledge and Beta doesn't take it as a reason to jump off a high ledge. Hence, several of the important features of *de se* thinking are absent in our case. Does this settle the question?

I don't think so. After all, these features are merely characteristic and need not occur in every instance of *de se* thinking. But, we can mount a further consideration that speaks against the *de se* nature of these inserted thoughts. It concerns the connection to the lived body. On the basis of experiencing her own body, Beta would most likely primitively self-ascribe the property of being *unable* to fly. She doesn't assume to be able to float around the room, safely jump from high buildings, and so on. The way she experiences the world around her through her body

speaks very much against her presumed ability to fly. In other words, the possibilities of interaction with the world are in conflict with the property she's supposedly self-ascribing in her believing (18). Hence, should we deny inserted thoughts the *de se* nature they might exhibit on the surface? I haven't and won't establish a conclusive answer. While some considerations speak against their being *de se*, it would require a much more detailed discussion of the phenomenon itself to evaluate the exact relation between inserted thoughts and the proposed account of *de se* thinking.

Finally, what about *de se* beliefs which involve the self-ascription of properties that the subject doesn't actually have? It might come as a surprise that this should pose a problem at all. After all, it's perfectly fine to have wrong *de se* beliefs. Many of our beliefs about ourselves are close to delusional but still remain beliefs about ourselves. Accordingly, if Alpha mistakenly thinks of herself as being generous, that amounts to a self-ascription of a property she doesn't have. No problem here. Somewhere in that process, an ascription to the lived body ensures that she's actually thinking about herself in the *de se* way. But the fact that she's self-ascribing a property she doesn't have doesn't impinge on the fact that her belief is *de se* in nature. Her belief being *de se* is grounded in her ability to primitively self-ascribe properties, but what she then believes about herself can be anything she dreams up.

What might be more troublesome are mistaken *de se* beliefs that are directly anchored in an ascription of a property to the lived body. For instance, imagine that Alpha doesn't actually have a horizontal back despite basing her belief (14) on her experiencing her back to be that way. Again, the problem sounds more serious than it is. The lived body isn't and doesn't need to be protected against epistemic errors. We witnessed this possibility already in the case of the rubber hand illusion. A subject might falsely assume that she can directly interact with the world through an artificial hand she accepts as her own. But such a misguided picture of the physical body doesn't undermine her capability to ascribe properties to the origin of egocentric space. It just results in *de se* beliefs which are potentially false. The lived body isn't supposed to be an absolute epistemic Archimedean point like Descartes's *Cogito*, it's merely the fallible basis for *de se* thinking.

This concludes the discussion of various kinds of *de se* thinking as explained by the idea that primitiveness is original thought. Subjects

are able to entertain *de se* mental states of various kinds because they ultimately think in an 'original' way: by ascribing properties to the lived body—the origin of their world. In this sense, all *de se* thinking involves a subject self-ascribing some property or other. But only some of these self-ascriptions are epistemically primitive and stand in direct connection to the lived body. For all other kinds, the *de se* nature of the belief is nonetheless fundamentally anchored in an ascription of a property to the lived body. For this is the foundation of a subject's ability to think about herself *as herself*.

4.4 FEATURE ACCOUNTING

After this important test of the aptitude of the lived body account, it's now necessary to look back to the characteristic features of *de se* thinking which were developed in chapter 1. Remember that these features are somewhat of a touchstone for any viable theory of *de se* thought since it's highly desirable that they're accounted for. All the other theories which have been discussed in the book so far failed in some respect to coherently accommodate all of these features into their account. Does the defended account, which traces self-ascription of properties back to ascription to the lived body, fare any better?

Since the account is a modification and advancement of the property theory advocated by Lewis, Chisholm, and Friends, we should expect similar merits and problems. After all, it still takes *de se* thinking to be tantamount to the self-ascription of some property or other. However, it differs in recognising the necessity of primitive self-ascription and the subsequent characterisation thereof. According to the defended account, the *de se* nature of every instance of thinking about oneself in the *de se* way ultimately relies on the ascription of a property to the *lived body*. This kind of ascription has to underlie every instance of *de se* thinking because it's needed to ensure that the subject can be aware of the fact that she's thinking about *herself*. As we'll see, this small, but fundamental, modification changes the explanation of the characteristic features quite radically.

So, let's start with the first feature: All *de se* mental states are about the thinking subject. In other words, whenever a subject engages in *de se* thinking, she's necessarily the intentional object of that thought. This semantic requirement is easily satisfied within the account. It was

argued that *de se* states of any kind are tantamount to self-ascriptions of properties. Because self-ascription is necessarily an ascription to oneself, a subject can only self-ascribe a property to herself. This results in the fact that *de se* beliefs are necessarily about the thinking subject, thus accommodating the first feature.

You might contest the trivial sounding claim that self-ascription is necessarily an ascription to oneself. The proposed account delivers good reasons that support this seemingly trivial fact. I've argued that self-ascriptions are ultimately anchored in the ascription of a property to the lived body. Since a subject can only ascribe properties to the lived body that constitutes herself, such primitive self-ascriptions have to be about the thinking subject. For instance, if Alpha believes that her back is horizontal, she ascribes the property of *having a horizontal back* to the origin. And since Alpha's origin is constituted through her lived body, that belief can't fail to be about herself. Hence, every primitive self-ascription takes the thinking subject as the intentional object. From this, we can infer that self-ascriptions which aren't primitive are also necessarily about the thinking subject. Because every self-ascription is anchored in the lived body, the account ensures that all self-ascriptions, by their very nature, are about the thinking subject.

The fact that self-ascriptions are necessarily about the thinking subject is closely connected to the second feature: The satisfaction conditions of *de se* thoughts depend in a systematic way on the thinking subject. For beliefs that amounts to the following: Whenever a subject entertains a belief of the type *I am F*, that particular belief will only be true if the thinking subject has the property of being *F*. Again, the idea that *de se* thinking amounts to the self-ascription of properties gives us an easy way to explain this second semantic feature. Every belief of the type *I am F* is an instance of a subject self-ascribing the property *F*. Whether or not such a self-ascription is correct naturally depends on who is doing the self-ascribing. If Alpha self-ascribes the property, her belief is true if Alpha is *F* because she's thereby ascribing that property to herself, i.e. Alpha. From the start, the concept of self-ascription is equipped with the required self-reflexivity. We can thus easily see how the satisfaction conditions of *de se* thoughts depend systematically on the thinking subject. It's because the correctness of self-ascription depends in a systematic way on which subject self-ascribes the property in question.

Things get more interesting as soon as we move to the epistemic features. Maybe the most discussed and important feature is that *de se* thinking is potentially immune to error through misidentification. That means that in some cases, a subject can't err about the object of her thinking. To put it differently, it sometimes isn't possible that she takes her belief to be about herself while it's potentially about someone else. We can compare this to cases where the intentional object—what the thought is supposed to be about—and the actual object of a mental state can come apart. To take an example, remember that Beta bases the belief concerning her own posture on seeing what she takes to be herself in the mirror. Hence, her belief relies epistemically on the identification of herself with the yogi in the mirror. Unfortunately, since there's the possibility that the mirror image doesn't originate from her, she might misidentify who she's thinking about. While the belief *My back is horizontal* is supposed to be about her—after all, that's the nature of *de se* thinking which is captured in the first feature—the indirect epistemic provenance of that belief entails the possibility of actually being about some other yogi.

We sided with Evans in claiming that the immunity which Beta is lacking has its origin in the fact that some of our mental states don't involve the identification of an object at all. Because the more experienced yogi Alpha doesn't identify herself when she bases her belief on her proprioceptive experience, she can't misidentify the object of her belief either. The result is that her *de se* belief is immune against that error. Why doesn't Alpha identify herself in the belief? Because she bases it on the epistemically direct ascription to the lived body. And I argued that subjects needn't identify the lived body because it's essentially first-personal. There's only ever one lived body through which a subject can interact with her world. And this epistemic uniqueness of the lived body makes an identification superfluous.

The possibility of immunity to error through misidentification is therefore explained by the nature of primitive self-ascription. Since subjects don't identify themselves as the object of their thinking in ascribing a property to the lived body, these self-ascriptions are free from identification. Therefore, *de se* beliefs are potentially immune to error through misidentification because they sometimes simply correspond to primitive self-ascriptions which don't involve identification of an object. And where there's no identification, there's no misidentific-

ation. Hence, we've established the possibility of *de se* mental states with immunity to error through misidentification.

This brings us to the intimate relation between *de se* thinking and self-knowledge. The presented account delivers an explanation of two different aspects of this connection. On the one hand, it explains how subjects know that their self-knowledge relates to themselves. And on the other hand, it unravels the possibility of substantial self-knowledge in the first place. Let's start by looking at the less problematic first aspect. Because knowing something implies believing something that's true, all cases of self-knowledge are cases of true *de se* beliefs. Furthermore, self-knowledge entails a subject's grasp of the fact that her knowledge concerns herself. The question now is: How can we account for this first-personal grasp? By intimately connecting *de se* thinking to the lived body. Accordingly, subjects are aware of their beliefs concerning themselves through the beliefs' being grounded in the ascription of a property to the lived body. Since subjects are the origins of their world, ascribing something to that origin involves an implicit awareness that you're thinking about yourself. And this grasp is brought to the surface in cases of self-knowledge. Thus, the structural peculiarity of self-knowledge is accounted for by how the account understands the nature of *de se* thinking. Self-knowing subjects are aware that they're thinking about themselves because every case of self-knowledge is a case of *de se* thinking, which involves the ascription of a property to the lived body.

The second aspect concerns the claim that self-knowledge is a special and superior kind of knowledge with special properties. Even if you doubt the truth of that claim, the proposed account could provide a potential justification of it. Primitive self-ascription was argued to be the distinguishing mark of *de se* thinking. Without the ability to self-ascribe properties in a primitive way, subjects wouldn't be able to grasp that they're thinking about themselves. Furthermore, the nature of primitive self-ascription is epistemically outstanding in being grounded in what a subject takes to be herself. It would be odd to inform Alpha that she's not actually in pain while she's clearly experiencing it through the lived body. This is because the ascriptions to the lived body are epistemically direct and through their immediacy less susceptible to epistemic errors. Thus, the claim that self-knowledge is somehow secured against certain doubts can be accommodated within our the-

ory because some items of self-knowledge have their source in this epistemically direct link to the lived body.

However, this result has to be qualified in two ways. First, it's purely conditional on the truth of the claim that self-knowledge is special at all. If we hold that primitive self-ascription is epistemically special, we're not claiming that it necessarily has to amount to self-knowledge understood in this superior way. We can recede to the much weaker claim that primitive self-ascriptions are a *better* source of knowledge than other self-ascriptions. Such a claim is much less ambitious and contentious. That some epistemic sources are superior to others is perfectly normal and needn't result in a distinction between different kinds of knowledge that correspond to these different sources.

Secondly, even if we accept that self-knowledge is a special kind of knowledge, and additionally accept that this is accounted for by the nature of primitive self-ascription, we can still allow primitive self-ascriptions to be fallible. While the pain example might tip us into the direction of embracing the infallibility of the lived body, the rubber hand illusion tells a very different story. Our primitive self-ascriptions are quite fallible indeed. But, as we saw in the first chapter, self-knowledge doesn't necessarily involve a claim of infallibility. We can be content with the claim that self-knowledge is special because primitive self-ascription is epistemically privileged. And this kind of privilege of self-knowledge might be all that's required for it to be special.

The fifth feature of *de se* thinking is the role it plays in motivating intentional action and producing behaviour. I argued that a subject has to take reasons that speak in favour of doing something as reasons for *herself* if they're supposed to be conducive to action. For instance, the fact that there's a charging honey badger to the left of Alpha is a reason for her to run the other way. However, that reason will only speak in favour of running away for Alpha if she believes *A charging honey badger is to my left*. In other words, it only motivates action if she takes it as a reason for herself by entertaining a *de se* belief. How does the idea that *de se* thinking is grounded in ascriptions to the origin of egocentric space account for this?

In fact, the lived body account tells a very harmonic and holistic story about this. Remember that we characterised the lived body via the assumed possibilities of interaction with the world. For instance, my fingers belong to the origin of my world because they're an assumed

means of experiencing and acting in the world. I can feel touch and warmth of the objects around me through my fingers and I can use them to type these words and sentences. But how does this transform anonymous reasons in the world into reasons for me?

Imagine that the letter 'R' on my keyboard stands in such and such a spatial relation to my left index finger. Accordingly, if I wanted to type that letter, I would have to move my finger in a certain way. In such a case, that particular spatial configuration would be a reason for me to move in a particular way. But, being a mere fact in the world, that reason is totally idle. The impersonal reason out there in the world is no business of mine. What is required is that I apprehend that reason in relation to me. How? By thinking about the letter in relation to my assumed possibilities of interaction—or: in relation to the lived body. Among other things, I ascribe the property of having an index finger in such and such a position—for example, bent and not stretched—to the lived body. This means that by thinking about the key from this original point of view, I take the relation between the finger, which I take to be mine, and the keyboard as a reason for me to act in some way or other. In other words, thinking about the key in egocentric terms provides it with immediate importance to my possible actions.

The proposed theory therefore provides a straightforward explanation of how intentional action is brought about by *de se* thinking. It traces our ability to think in the *de se* way back to how we think about our world from its origin. Subjects think about the world around them in an egocentric way that's essentially tied to the lived body. The lived body, in turn, is characterised through the assumed possibilities of interaction. This naturally makes certain facts in the world connect with my potential actions and thus produces potential reasons for me to act. Thereby, these conceptual and epistemic links provide a simple explanation of this last feature of *de se* thinking.

We can thus see that the lived body account is capable of doing justice to all the characteristic features of *de se* thinking that we've identified. Hence, it passes the second important test for any theory of *de se* thinking. Not only is there a reasonable explanation of various kinds of *de se* thoughts in terms of ascription to the lived body. The account also ensures that *de se* thinking is always about the thinking subject, comes with satisfaction conditions that systematically depend on who's thinking, accounts for the possibility of immunity to error

through misidentification and the possibility of self-knowledge, and explains the conceptual link to intentional action and behaviour.

4.5 ANSWERING OBJECTIONS

Of course, philosophy isn't so easy and straightforward. The proposed account might fit in fair and square with various desirable features and examples but it certainly isn't immune to objections. Some of these have been discussed and accounted for in the process of developing and explaining the account in the first place. But this still leaves us with the rest. Let me now answer some of the more important and potentially devastating ones before wrapping this book up.

The first objection concerns the way the lived body is characterised. The account of *de se* thinking which I defended is very much tied to a subject's assumed possibilities of interaction with the world. According to this phenomenological theory, the lived body—the source of all *de se* thinking—is established through the way a subject perceives and acts in her world. But what if these possibilities vanish? What if a subject is incapable of experiencing her own body and the world around her and incapable of any bodily movement? Does her capacity to think in the *de se* way disappear as well?

Christopher Peacocke presents us with one version of this objection. He imagines a subject in such a predicament. She has no proprioception and no way of moving her body. Presumably, she assumes no possibilities of interaction in that situation of sensory deprivation and impossibility of action. Accordingly, her lived body should evaporate completely. And without the possibility of engaging with the world from the origin, all *de se* thinking should vanish. However, Peacocke, following similar remarks by Elizabeth Anscombe (1975), thinks it doesn't: 'But this subject's use of *I* still refers in these circumstances. … [T]he subject may think to herself *I will make sure I don't get into this situation again*' (Peacocke 2012b: 153).

Similarly, Lucy O'Brien argues that the absence of bodily awareness, which partly constitutes a subject's lived body, doesn't result in the impossibility of *de se* thinking. Like Anscombe and Peacocke, she assumes that a subject can still think about herself even if she's in a state of sensory deprivation: 'Now, even the subject immersed in the sensory deprivation tank is able to refer to herself first-personally. So,

however the subject is presented with herself, it cannot be via those perceptual sources that are unavailable to her in such a situation' (O'Brien 2007: 20).

In both versions of this objection we have the claim that a subject who's incapable of bodily experience and behaviour, but who's still in her right mind, is capable of entertaining *de se* thoughts. This is a problem for the defended account insofar as the subject doesn't assume any possibilities of interaction with the world and would therefore be devoid of a lived body. And without a lived body, there are no primitive self-ascriptions and thus supposedly no *de se* thoughts. Hence, it seems that the account implies that a subject in sensory deprivation can't think about herself in the *de se* way, while the example suggests it to be possible.

To be sure, the reasons why Peacocke, Anscombe, and O'Brien assume this scenario to be an actual possibility are somewhat different, but we don't need to discuss them individually to dismiss the objection. We'll just accept that such a scenario is possibly. Presumably, the subject we're talking about is a highly self-conscious and sophisticated subject. As such, she's capable of correctly using the first-person pronoun and thereby expressing thoughts involving the correct application of the first person concept. This already opens up a crack wide enough to drive in an argumentative wedge.

The easy way to dispatch the objection is to argue that a subject who's still capable of linguistically expressing her thoughts has to assume that she's capable of uttering words. In other words, she assumes that she can produce sound by moving her lips and vocal chords in a certain way. As such, she still retains a heavily constricted lived body to which she can ascribe properties. Hence, if we allow our subject to be capable of language, we at once bring back the lived body. But this reply is much too easy. This is because we can interpret the objections to exclude the possibility of linguistic expression. Rather, the subject is so heavily immobilised and desensitised that her thoughts are all that remains.

This makes it much harder for the lived body to get into play. For the sake of the objection, we may assume that the lived body is truly gone in this particular scenario. Does that eliminate the possibility of *de se* thinking? Remember that the approach I defended assumes that every instance of *de se* thinking involves an ascription to the lived body because this is what establishes a subject's grasp of the fact that

she's thinking about herself. As a reply to the objection, we can specify this requirement by adding that this kind of involvement need not be contemporaneous. Rather, a subject's *de se* thoughts can be based on a prior ability to ascribe a property to the lived body. This allows the subject to retain the epistemic base on which her first person concept is built. How so?

First off, a subject can't acquire the first person concept without the ability to primitively self-ascribe properties. Without the general capacity to think in the *de se* way, she wouldn't understand what it means for the first person concept to refer to *herself* when she's using it. Thus, our sensory deprived subject used to be capable of ascribing properties to the lived body. Otherwise, she wouldn't now be able to think thoughts involving the first person concept. This follows from our discussion of the conceptual strategy. In her predicament, she can now epistemically base her use of the concept on that prior ability to ascribe properties to the lived body by simply applying the concept in question. Her full grasp of the concept already involves grasp of the fact that her use of the concept refers to herself. Absent any impairment which would destroy that grasp, she retains mastery of the concept. Her application of the first person concept in the sensory deprivation scenario thus epistemically involves a prior ability to ascribe properties to the lived body. This still stays true to the idea that all *de se* thinking epistemically—not temporarily—involves an ascription to the lived body.

Thus, the subject—being self-conscious—retains her mastery of the first person concept despite the temporary 'disappearance' of the lived body in the sensory deprivation scenario. As such, she retains her ability to use that concept to self-consciously refer to herself in *de se* thinking. That ability, in turn, is based on primitive self-ascriptions which epistemically buttressed her acquisition of the concept in the first place. This is why subjects can think *de se* thoughts even if they're currently incapable of ascribing properties to the lived body. Hence, the possibility of *de se* thinking in a sensory deprivation scenario is compatible with the lived body account.

A second objection tries to undermine the claim that primitive self-ascriptions can be immune to error through misidentification. It argues that primitive self-ascriptions simply don't have the right structure to be candidates for immunity. Why not? Because the logical and semantic structure of primitive self-ascription doesn't involve an object

which could potentially be misidentified. Hence, it's wrong to say that ascriptions to the lived body can be immune to error through misidentification in the same way that it's wrong to say that clouds are immune against chickenpox. It's a simple categorical mistake.

The idea behind this claim is that some of our mental states have a structure which distinguishes between the object that's picked out and the property that's ascribed to that object. Other mental states don't involve such a structural distinction; rather, the information about the object is merely implied. For instance, there's a clear object and a clear property in a subject's belief that politicians should be honest. She's thinking about all the politicians out there and she wants them to have the property of being honest. But, we might describe the visual experience of drinkable water in front of me as a mere positing of drinkable stuff in my visual field. Such positing would only involve a property at a certain location in egocentric space without a clearly designated object having that property (A.4.13).

Kristina Musholt formulates this point in the context of the distinction between non-conceptual and conceptual thought. We can admit that some of our self-ascriptions aren't based on the application of concepts. As such, she argues, they don't involve the explicit individuation of an object to which a state or property is ascribed. Rather, the object of one's mental state is implicitly provided by the nature of the state in question. Being a self-ascription, the property is naturally ascribed to oneself. Thus, the fact that we're dealing with a self-ascription of a specific subject delivers us the right object without that object featuring in the self-ascription. She gives the example of feeling irritated because of one's hunger. Because a subject can't self-ascribe the property of being irritated to any other object than herself, 'such a self-ascription can "what"-misrepresent, but it cannot "who"-misrepresent' (Musholt 2015: 69). In other words, self-ascriptions don't have a subject–object structure. And immunity only applies to states with such a logical structure.

She takes this analysis of self-ascriptions—which we can take to apply to primitive self-ascriptions as well—to speak against the possibility of immunity of these states. Obviously, where there's no object involved, no object can be misidentified. As a result, *de se* states which amount to ascriptions to the lived body are not within the realm of misidentification at all. Hence, they also can't be immune against such

an error. The question of identification simply doesn't arise on that level. For Musholt, this is why 'it is a category mistake, so to speak, to try to apply the notion of immunity at the level of nonconceptual content' (Musholt 2015: 70). If the objection goes through, we would have to revisit the claim that the lived body account can accommodate the possibility of immunity.

The general idea behind this objection can also be found in other philosophers who write about immunity. For instance, Peacocke applies the notion of immunity only to conceptual judgements in which an object is identified: 'The relevant notion is that of a judgement with the content *Fa* being immune to error through misidentification (a) relative to the particular occurrence of *a* in the content, (b) when the judgement is reached in a certain way W, and (c) in normal circumstances' (Peacocke 2014: 107). Hence, immunity simply doesn't figure on the level of primitive self-ascriptions because they don't amount to conceptual judgements. How can the lived body account react to this?

I propose two different strategies and will defend one of them. The first one is much more difficult to justify and attempts to question the restriction of immunity to conceptual judgements where an object is clearly identified in thought. We can reasonably ask why the concept of immunity shouldn't apply to those mental states where the object is merely implicitly present. Self-ascriptions might not make explicit reference to the object, but it certainly involves a clear reference to some object or other. After all, the object to which the property is ascribed is delivered to us by the fact that it's the self-ascription of a specific subject. In other words, such a *de se* state isn't mute about who the property is being ascribed to. Hence, there's at least the possibility of something going epistemically wrong in the process of ascription to the lived body. And consequently, there's a sense in which we can claim those self-ascriptions to be immune to error through misidentification.

This strategy might be viable but it would require a more detailed discussion of immunity in order to get off the ground. What's more important is that we're not forced to go that way. The lived body account makes no substantial claims about immunity. It only needs to be compatible with the possibility of immunity of *de se* thoughts in general and explain how the phenomenon can arise. The good thing is that this can be done while accepting the objection that the notion of immunity can only be applied to conceptual judgements.

Here, then, is the second strategy to defuse the objection. We can argue that the immunity of judgements is precisely accounted for in terms of the lived body. The reason why some first person judgements are immune to error through misidentification is that they're grounded in a primitive self-ascription. The way, to use Peacocke's term, in which the judgement is reached provides an epistemic foundation which is secure in a way that makes the judgement established thereupon immune to error through misidentificaion. We can't ascribe a property to the lived body and err about who we're ascribing that property to. And this fact—which might not imply that these primitive self-ascriptions themselves exhibit immunity—epistemically secures that judgements formed on that particular basis are immune to misidentification. In fact, Peacocke (2014: 107) provides us with a short list of characteristicly immune judgements which are all ultimately formed on the basis of an ascription to the lived body. Hence, the lived body account can also accommodate this restricted notion of immunity.

The third and final objection I want to discuss is somewhat more global and concerns the general strategy of the lived body account. It goes back to Descartes's attempt to secure all knowledge by basing it on self-knowledge and therefore on some form of *de se* thinking. The famous *Cogito* argument can be understood as the establishment of an absolutely secure form of knowledge. And this kind of self-knowledge is so epistemically immaculate because it's completely independent of any knowledge of the outside world—including anything like a body. A disembodied spirit could go through the Cartesian Meditations just as well as you and would arrive at the same conclusions.

Given the influence of this idea—the knowledge of one's own existence is probably as close to an undisputed philosophical fact as we can get—there's an obvious tension to the lived body approach. The epitome of self-knowledge is supposedly knowledge that's completely epistemically independent of a subject having a body at all. In addition, the Cartesian strategy is designed to explicitly annul all reference to a dubitable and fallible empirical or phenomenological body in its quest to build a foundation for knowledge. How can the lived body account react to this general and potentially destructive objection?

One seemingly fruitless reply uses the strategy against the first objection to mount a defence of the necessity of primitive self-ascription for *de se* thinking. According to this defence, the Cartesian argument

to arrive at knowledge of one's own existence only works for a subject that's already capable of *de se* thinking. And this would imply that the subject comes to understand that her first person thoughts are about herself by basing them on ascriptions to the lived body. However, this strategy doesn't seem as compelling against Descartes's worry. We can interpret the *Cogito* as establishing that some form of self-knowledge is completely independent of a subject having a physical body at all. And this is a more serious predicament than the sensorily deprived subject in O'Brien's scenario. While it's possible to argue that a subject's mastery of the first person concept is epistemically grounded in some prior primitive self-ascription in the latter case, this strategy seems unavailable to us in the Cartesian scheme.

We need to look for another option then. One attractive reply is to question the intelligibility of Descartes's scenario in order to make the previous reply viable. As such, it simply doesn't accept the premise that a disembodied subject can achieve knowledge of her own existence. To be more precise, this strategy stresses that the general possibility of self-knowledge only arises once a subject grasps that some of her thoughts are about *herself*. And this requires the ascription of a property to the lived body. For otherwise, how could the subject grasp that she's thinking about herself? While the specific item of self-knowledge that's established in the *Cogito* argument is epistemically independent of any particular empirical or phenomenological premise concerning the body, her *ability* to think about herself in the first place isn't independent. As such, while we can imagine ourselves going through the Meditations and arriving at that profound piece of self-knowledge, a disembodied spirit can't follow suit. Without a lived body, the spirit can't come to understand what it means to think about oneself in the first place.

How plausible is this reply? First off, one shouldn't mistake the reply as going against the metaphysical theory of dualism—the idea that there are independent mental things next to physical things. The claim is merely that a pure spirit can't arrive at self-knowledge of any kind because it can't primitively self-ascribe properties. And therefore, it can't entertain *de se* beliefs at all. Secondly, the reply somewhat begs the question in resulting in the mere claim that the Cartesian scenario of a disembodied spirit with self-knowledge, which we deemed possible, is actually impossible. Apart from the theory of primitive self-

ascription itself, it doesn't provide additional reasons that challenge the possibility of disembodied spirits with self-knowledge. And this is certainly problematic.

However, I argue that it's the best we can do and it's the best we *need* do. Descartes doesn't provide us with any reasons for accepting that particular possibility. In fact, there are countless other reasons speaking against it. So, why should we be under pressure then? From the fact that the *Cogito* argument is epistemically independent of any empirical knowledge, it doesn't follow that disembodied spirits can think *de se* thoughts or that such things are even possible. In fact, the important point of the *Cogito* argument can be accommodated neatly into the lived body account. As ordinary subjects with lived bodies, we can arrive at self-knowledge that's epistemically independent from any particular empirical ascription to the lived body, but this depends on the general ability to ascribe properties to the lived body. This central result of the Meditation is totally independent of the further claim that pure spirits can have the same kind of self-knowledge. And as such, we can remain mute. The knowledge of one's own existence is thus just as dependent on the lived body as the subject's capacity to think in the *de se* way while sensorily deprived.

4.6 CLOSURE AND LOOSE ENDS

We've now answered some critical objections and thereby further consolidated the lived body account. This only leaves a few small tasks that need to be accomplished. The first of these is a mission of peace. In the course of the book, several opposing strategies were critically discussed and dismissed. Most of these dismissals were motivated by the following consideration: While they provide an explanation of how established *de se* thinking works semantically or epistemically, they all fail to explain how subjects can come to grasp that their *de se* thoughts are about themselves in the first place. In other words, the main objection was that they fail to give an account of what's actually special about *de se* thinking. But this shortcoming doesn't make the accounts wrong, they're just incomplete.

In theory, then, the lived body account can be seen as an attempt to complement and complete these very different accounts by explaining the primitive nature of *de se* thinking. The essential elements of the ac-

count can be transplanted into the linguistic, the conceptual, and also the property theory to discharge their duty of explaining the possibility of *de se* thinking. Obviously, how exactly this is to be done remains the job of proponents of these theories. Nonetheless, we can give some hints as to how this might be achieved in order to kickstart such attempts. However, it's important to emphasise that the theory which I've defended is intended to be complete and independent of the other discussed alternatives. As such, it's possible to accept it without feeling the need to incorporate the other strategies.

First off, let's take a look at the closest relative to the lived body account: the property theory defended by Lewis, Chisholm, and Friends. The main problem was their failure to clearly distinguish proper *de se* thinking from other kinds of thinking. More importantly, I argued that the theory has to accept some form of primitive self-ascription as a basis of genuine *de se* thought. Unfortunately, no explanation of this necessary kind of self-ascription was provided. The analysis of primitive self-ascription as ascription to the lived body can be used to fill this gap in the property theory. For instance, this could yield the more precise claim that what it means for a subject to ascribe a property under the relation of identity is that she ascribes the property to the lived body—the thing which constitutes her as a subject. Now, I don't want to claim that defenders of the property theory would accept this proposal. In fact, the strong phenomenological coating of the lived body might not fit in well with the underlying functionalism that's advocated by David Lewis, to give but one example of a potential conflict.

Next up, the conceptual theory was illuminated through Peacocke's work. Its essential part is the idea that *de se* thinking can be explained through a subject's knowing use of the first person concept. As indicated, Peacocke himself accepts a more primitive first-personal notion that underlies mastery of the first person concept. Unfortunately, the exact nature of this *i* notion remains somewhat vague and unclear. An obvious way to reconcile the two quite distinct accounts would be to interpret the lived body account as an illumination of that non-conceptual notion. After all, the replies to some of the objections against the defended theory made the relation between primitive self-ascription—understood as ascription to the lived body—and the acquisition of the first person concept quite explicit and clear. That concept can only be mastered by anchoring it in the ascription of a property to the lived

body. This allows the subject to grasp that her use of the concept applies to herself. At the same time, we saw in the replies to the objections that once such a concept is learned, it can operate quite independently from primitive self-ascriptions. Therefore, the two quite distinct ideas and accounts could potentially be merged.

Closely related to this is the linguistic approach. Again, we saw the necessity to explain how a subject understands that her use of the first-person pronoun refers to herself in speech. Primitive self-ascription provides a very elegant way to do this. A quote from O'Brien wonderfully clarifies the support that the lived body account gives to the linguistic approach:

> We need more than knowledge that 'I' refers to its producer in order for it to be the case that I know that a particular token of 'I' refers to me. (…) The missing element was provided by the fact that I have a certain kind of awareness of my actions. (…) The awareness of my actions, that I create by acting, thinking and uttering, is of a different and more primitive kind than the self-awareness constituted by a capacity for first-person reference.
>
> O'Brien 2007: 78–79

In other words, by linguistically engaging and interacting with the world through the lived body, the subject becomes aware of her own use of language. As she moves her lips, she becomes aware of being the producer of that sound and thus, according to the meaning rule, its reference. This delivers the key to understanding that her own utterance of the first-person pronoun—being governed by some linguistic rule—refers to herself. As such, the engagement with the world through the lived body anchors a subject's use of language.

This brings us to the end of our quick tour of alternative approaches and simultaneously to the conclusion of this book. Time to recapitulate the various stages and insights that were developed. We started our journey with the double-edged sword of Narcissus's self-consciousness. On the one hand, self-consciousness is a highly sophisticated and marvellous form of thinking. It allows us to build complex societies, laws, write interesting books, or fly to space. On the other hand, it unravels the deep mystery of our place in the universe. By dwelling on the

nature of our own existence, we come to explore the uncertain meaning of life and our own purpose. A conundrum that is out of reach for subjects without self-consciousness.

But what's special about the ability to think about oneself? Exploring this question, we saw that several semantic and epistemic features are characteristic of *de se* thinking. By comparing it to other ways of thinking about oneself, such as thinking *de dicto* or *de re*, we saw that only this intimate kind of thinking about oneself is necessarily about the subject. But we also explored the connections to epistemic aspects of thinking. In some cases, subjects can't err about the object they're actually thinking of. Their *de se* thoughts are immune to misidentification. And, of course, self-knowledge, that profound insight brought to perfection in Descartes's *Cogito*, is intimately tied to our ability to think in the *de se* way. Finally, and most controversially, our actions are only motivated insofar as we apprehend reasons in the world as *our* reasons. And that again requires us to take a *de se* stance towards them.

With these features unearthed, we were ready to explore some options of explaining the nature of *de se* thinking. We followed the analytical tradition and looked at the Propophile strategy. Here, we discovered an important problem: The way we think about ourselves is highly dependent on the context and can't be easily captured by a rigid tool such as a proposition—a possible way the world could be. Instead, we have to build a multi-layered semantic building which can account for the flexibility of *de se* thinking. While this was quite successful in dealing with the semantic features, we exposed a fundamental and fatal flaw. All these two-dimensional strategies couldn't explain how subjects grasp that their thoughts are about *themselves*. They explained how reflexive self-reference worked, but they didn't explain the underlying epistemic nature of self-conscious self-reference.

As a result, we traced our steps back and opted to go into a different direction. The lesson of the Propophile failure isn't to make the propositional picture more complex and encompassing but to relinquish it and break new ground. We discussed the omniscience of goddesses and explored the property theory defended by Lewis and others. In that context, we realised the potential of the seemingly trivial claim that *de se* thinking just amounts to self-ascription of properties. However, I simultaneously manoeuvred the theory into a potentially devastating impasse. It ran the risk of utterly blurring the line between

de se and *de re* thinking—an unacceptable outcome. I argued that we need to postulate a primitive kind of self-ascription that isn't open to the analytical attempts of reduction. Only such primitive self-ascription can really explain how subjects grasp that they're thinking about themselves.

This move back to the primitive was the steppingstone to develop a theory of primitive self-ascription that's capable of completing that crucial task. I borrowed the concept of the lived body from phenomenological insights about how subjects engage with the world around them. As origins of egocentric space, subjects apprehend themselves as world-engaging subjects. And the lived body is the means through which this engagement manifests itself. Primitive self-ascription was then traced back to ascription to the lived body. Because the lived body is given to the subject in an immediate and essentially first-personal way, these ascriptions are fit to play the grounding role for *de se* thinking. And the rest is history. I examined how this approach deals with several typical examples of *de se* thinking, how it accounts for all five characteristic features of *de se* thinking, and defended it against potentially devastating objections.

Of course, the jury is still out on whether such an approach is completely viable and consistent. Moreover, some interesting questions and fields of research have been left untouched. Before bringing down the curtain, let me explore a few of these uncharted realms. I shall start with the question that originally motivated me to delve into the topic of *de se* thinking: Can nonhuman animals be self-conscious? It's a well-known fact that some individuals of some species—such as chimpanzees, elephants, or dolphins—can pass the mirror rouge test. They can recognise themselves in a mirror and interact with their mirror image in very much the same way as humans do. But what exactly does it mean to pass that test? What does it tell us about the cognitive capacities of these subjects?

The topic of animal minds is a very broad and convoluted one. Many different interests and methodologies are released on a seemingly homogenous thing. This creates interesting results but also a lot of disagreement and conceptual muddle. Nothing guarantees that the neuroscientist uses the concept of self-consciousness in her research in the same way as the philosopher and the cognitive anthropologist. Furthermore, it's quite unclear whether the minds of animals are a monolithic

kind—requiring just a single key to be unlocked. Consequently, all these results have to be assessed very carefully and combined into a coherent picture. And this is an enormous task which is still undone. Maybe the theory of the lived body which was developed here can shed some light on some of the results and questions concerning self-consciousness in nonhuman animals. After all, the theory is supposed to have a threshold low enough to allow for *de se* thinking even in very simple subjects and organisms (A.4.14).

This leads us straight to another current field of research: artificial intelligence. Many philosophers and other researchers are extremely reluctant to grant consciousness and self-consciousness to current artificially intelligent beings. But how exactly do we need to understand the seemingly cognitive processes underlying the behaviour of robots, autonomous drones, and complex machines? Many of them are capable of learning according to complicated algorithms. But what distinguishes this kind of learning from human and nonhuman animal learning? Artificial intelligence is an interesting test case where we can see how far we can push our conceptual theories.

The lived body account is heavily based on phenomenological data such as the experience of one's engagement with the world around us. We, as conscious beings, have a good grasp of how this works in human beings. But can we transplant this to the field of artificial intelligence? It seems that the way self-driving cars represent their environment is very akin to how we think egocentrically about the world. What prevents us from saying that these cars are capable of *de se* thought? Their supposed lack of consciousness? Might it be most parsimonious to grant them the ability to think in the *de se* way?

Finally, it's important to discuss a more conceptual shortcoming which isn't dependent on empirical uncertainties or science fiction scenarios. It's the simple fact that the discussion about *de se* thinking is almost exclusively framed within the context of cognitive states like beliefs or perceptions. In these cases, it makes sense to use talk of ascriptions. But can we easily extend this to conative states such as intentions, desires or emotions like shame and pride? The belief that I'm tall is accurately represented by claiming that I self-ascribe the property of being tall. However, which property do I self-ascribe when I intend to go hiking? Which primitive self-ascription underlies an emotion?

Many philosophical theories about the mind are supposed to be general and encompassing. As such, they're intended to cover not just states like judgements and beliefs—states which aim to correctly picture how the world actually is—but also states that are directed at how the world should be. The lived body account is no exception to this ideal. However, the exact way to accommodate desires, intentions, or emotions is not immediately clear and remains to be determined. While I would argue that it's possible to extend the idea that *de se* thinking is grounded in the lived body to these states, I don't present a clear analysis or strategy of how this is fleshed out.

This concludes our short survey of yet to be charted scientific and conceptual realms. As Herman Melville's Ishmael so impressively conveys, a complete account of anything is neither within the scope nor the possibility of any human endeavour. This book is no exception. It's but a small step into the direction of a more complete, accurate, and encompassing theory of how we think about ourselves. Nonetheless, I hope that some important messages and points have been made as clear and convincing as possible. What remains is my invitation to you to develop the story further and navigate the insidious seas of conceptual thought. ■

A

AUXILIARIES & TECHNICALITIES

A.1 ADDENDA TO CHAPTER 1: BEING IN THE MIRROR

A.1.1 *The relation between de re and de se thinking*

The distinction between the three kinds of thinking about oneself can and needs to be clarified further. First off, the primary contrast is between thinking *de re* and thinking *de dicto*. Sainsbury and Tye (2012) point to the difference between believing *The murderer—whoever she may be—has large feet* and believing *That woman has large feet*, while pointing to a particular woman in sight. They argue that the latter is a *de re* thought because it can only be properly used if the thinking subject knows who the woman is. In contrast, the former can be thought 'not knowing who the murderer is' (Sainsbury and Tye 2012: 122). So, thinking *de re* entails some direct reference in thinking to a specific particular object. However, there's significant disagreement concerning the nature of this direct reference and how it's mediated. For instance, Sainsbury and Tye argue that *de re* attitudes are those attitudes in which a *de re* concept forms part of the content. Burge (2007), on the other hand, seems to require that at least some of the content isn't conceptualised.

Now, the case of *de se* thinking complicates matters slightly. This is because it's similar to *de re* thinking in a very important sense. Like *de re* thinking, it involves the direct reference to a particular object which isn't necessarily mediated by a concept (cf. Burge 2007: 68). The successful reference to a specific thing is instead mediated by demonstrative, indexical, or other directly referential elements. So, there is an important resemblance between the two kinds of thinking. However,

the exact relation between *de re* and *de se* attitudes isn't determined and is open for discussion. It heavily depends on the specific scope which is targeted—e.g. whether we're talking about the third-person ascription, the epistemic foundation, or the semantic content. This disparity results in Lewis (1979) claiming that the *de se* subsumes the *de re*, while Perry (1980) thinks it's the other way round. Still others, like Récanati (2009), hold that it depends on the exact nature of the *de se* attitude in question: sometimes *de se* believing is also *de re* and sometimes it isn't.

For the purposes of this book and the arguments therein we can settle on the following claim concerning the relation between *de re* and *de se* attitudes: Insofar as *de re* thoughts are mainly characterised by referring to a specific particular thing—where this reference is unmediated by some conceptual or descriptive element—*de se* thoughts are types of *de re* thoughts. Nonetheless, like *de dicto* attitudes, the reference of *de se* beliefs is sensitive to the context and can change from one possible world to another.

A.1.2 *What exactly is an intentional object?*

The concept of an intentional object derives from the idea that intentionality is the basis of mental states. This is especially prominent in Brentano (2009) and Husserl (2001) where intentionality is characterised as the 'aboutness' of mental phenomena. Normally, when a subject believes, desires, or hopes something her belief, desire, or hope are *about* something. It's about the future of the world, where she hopes that peace will finally prevail. It's about the solution of a mathematical problem, where she believes that it's $\int_0^\infty e^{-x}\,dx$.

The famous quotation from Brentano's *Psychology from an Empirical Standpoint* introduces the concept of intentionality in a livid way:

> Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomenon includes something as object within itself, although they do not all do so in the same way. In

> presentation something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired and so on.
>
> Brentano 2009: 68

In the case of *de se* thinking we can say that the thinking subject always features in the intentional object of the thought. When Alpha thinks *I am F* she thereby thinks about herself. And hence she's the intentional object of her thought. We can generalise this insight into the following theorem: For every subject S entertaining an instance of *de se* thinking θ, S is the intentional object ω of θ.

However, this doesn't imply that the subject of *de se* thinking is the intentional object in the same way as the apple is the intentional object of Adam's desire. There might be different ways of thinking about something. In some cases a specific thing might need to be identified and picked out from a number of possible intentional objects. So, a subject might think about that particular red ball in snooker as opposed to the other red ball next to it. And there are also different ways of taking oneself as the intentional object—even within different kinds of *de se* thinking. In some case, the subject is the object of her thinking in a very reflective way as when she's introspectively thinking about her character traits or how she treated a friend in a discussion. In other cases no such reflection is needed and the intentional object of one's *de se* thinking is determined pre-reflectively in virtue of the *de se* nature of that particular intentional state.

Furthermore, we have to distinguish different ways of procuring the intentional object of a *de se* attitude. If a subject mistakenly believes that she's seeing her own reflection in the mirror and bases her belief *I am looking tired* on that visual experience, the intentional object of that belief is still herself even if she misidentified herself in thinking. The claim that every instance of *de se* thinking entails that the thinking subject is also the intentional object isn't the same as claiming that all *de se* attitudes are immune to error through misidentification or have referential security. The former is true while the latter is obviously false.

A.1.3 *Truth-value of propositions*

We can put the Propophile's original statement into more formal terms by saying that for every proposition p and for every possible world *w* the truth-value of p at *w* is determined. Or, in other words: for every proposition p and for every possible world *w*, if p is true at *w*, then it's not possible that p is false at *w*.

This is of course not accepted by everyone. Depending on how exactly you characterise either of the relevant terms in the definition above—truth, proposition, possible world, truth at or in a world—you can develop different claims about the truth-values of propositions. For present purposes we stick with the Propophile characterisation because it's easily understood and closely resembles classic theories about propositions. Thus, it's suitable to illustrate their inherent problems with *de se* thinking.

A.1.4 *More on propositions*

The relation between a believing subject and the proposition believed is another contentious matter. The standard Propophile way of putting it is to say that a belief amounts to a relation of entertaining between a subject and a proposition. This is of course mightily vague, but I won't comment further on that.

What's more important to elaborate on is the relation between the idea of belief, understood as the entertaining of a proposition, and the concept of an intentional object. Some beliefs are properly about propositions but we don't *always* want to say that the intentional object of a belief—what a belief is properly about—is a proposition. If I believe that Sydney is the capital of Australia, my belief is about the city Sydney and not about some proposition or other. Soames (2014) suggests that propositions can't do the job the Propophiles signed them up for. Instead, propositions are useful devices for theoretical purposes. We can use them to categorise subjects' beliefs according to types. Thus, a proposition is a type of cognitive event. According to this idea propositions are 'pieces of information that represent things in the world as being certain ways; thus they have truth conditions. Since the proposition that o is red represents a certain object as red (while doing no further representing) it is true iff o is the way it is represented to

be—red'. And when a subject believes something she *entertains* a certain proposition. And that again is to do something particular—thus propositions are types of *events*—e.g. 'to entertain the proposition that o is red is to *predicate* redness of o, and thereby to *represent* it as red'. (Soames 2014: 95)

This kind of reasoning undermines the conception of a proposition 'out there' in abstract reality to which subjects 'hook up' to when they entertain a proposition—i.e. believe something. In fact, such a picture is far from trivial. And this fact can be established on grounds which are quite independent of the problem of *de se* thinking. However, even Soames's notion of a proposition is problematic—for very much the same reasons as the Propophile's propositions will prove inadequate. In *de se* thinking a subject doesn't necessarily *represent* the world in a way which directly delivers conditions of satisfaction. Rather, we need some kind of indirect two-dimensional treatment of propositions if they are to work for *de se* thinking as well. Soames's theory doesn't preclude such a treatment, but it also doesn't entail it.

A.1.5 *Mental and physical properties*

We need to be a bit careful when we're dealing with the nature of mental and other typically first-personal properties such as *wishing*, *feeling*, *contemplating*, or *remembering*. Typically, I know best when I'm remembering something or not; maybe because a remembrance has a distinctive cognitive 'feeling'. But, of course, it can be determined by outsiders as well whether I'm remembering something or not. So, we often say things like 'Ah, now she's remembering what happened last night'. However, the way other people come to this conclusion is quite different from the way that we come to the conclusion that we've just remembered something. While others have to rely on a certain form of evidence—what the subject says, how she behaves, and so on—we can know directly from the first-person perspective whether we're remembering something or not. This is especially prominent in cases of *déjà vu*.

However, the case of déjà vu also shows that this difference has nothing to do with the question whether our introspective beliefs are always correct. We can be wrong with respect to the mental properties we ascribe to ourselves just as others can be wrong about these things.

But in the first-personal case the epistemic route can be direct whereas it's always indirect in the third-personal case. So, the important point is that many mental properties can be ascribed to oneself on an epistemic basis that is qualitatively distinct from the way we know about the mental and bodily properties of others.

Perry (1990, 2002) accounts for this feature by explaining that many mental properties can be ascribed on the basis of what he calls 'self-informative ways of knowing'. Here's how he defines this special kind of knowing:

> A perceptual state S is a normally self-informative way of knowing that one is ϕ if the fact that a person is in state S normally carries the information that the person in state S is ϕ and normally does not carry the information that any other person is ϕ.
>
> Perry 2002: 204

So, this way of knowing is cognitively built in such a way that it guarantees to carry information from the very same agent that receives and processes this information. In other words, this architecture ensures that the beliefs formed on the basis of this information can't be about anyone else but the thinking subject. Since our beliefs about the mental properties of others don't have this cognitive architecture they aren't immune to error through misidentification relative to the ascription of a mental property. Yet, they might still be demonstratively immune to error through misidentification.

A.1.6 *Does the Cogito result in self-knowledge?*

There's a long standing debate concerning the conclusion of Descartes's *Cogito* argument. Lichtenberg (1990: 168) proposed that the only thing we can know on the basis of our introspective awareness of thinking, doubting, or wondering, is *There is thinking*. In other words, we can't reach self-knowledge on the basis of the *Cogito*. We only achieve knowledge of the fact that some instance of thinking is going on. Campbell (2012) illustrates the reasoning for this conclusion:

> Let's go back now to the *cogito*. Can it be regarded as explaining how you know of your own existence? The picture I would recommend here is that knowledge of your

> own existence is already required by the transition from (1) having a particular conscious thought, to (2) knowledge that you are thinking. The argument is that regarding the mere having of a conscious thought, as grounding knowledge of the judgement that one is thinking, already presupposes that one exists. The transition from (1) to (2), therefore, cannot be thought of as grounding or explaining one's knowledge of one's own existence.
>
> Campbell 2012: 364–365

If you want to derive knowledge of the fact that *you* are thinking from the awareness of a conscious thought, you have to already have knowledge of *your* existence. He compares this to Moore's famous 'proof' of the external world. In this case the visual experience of your own hands is intended to support knowledge of the fact that there are two hands: 'How? By holding up my two hands, and saying, as I make a certain gesture with the right hand, "Here is one hand", and adding, as I make a certain gesture with the left, "and here is another"' (G. E. Moore 1939: 296). Many think that this common sense argument doesn't work as it simply begs the question (Coliva 2003; Pryor 2004). Knowledge of the existence of external objects is presupposed as one moves from the visual experience to the judgement that there are two hands. Supposedly, something similar applies in the case of the *Cogito*.

In a direct exchange with Campbell, Peacocke (2012a) defends the Cartesian conclusion. He argues that the disputed move in the argument is justified by 'plausible conceptions of consciousness, the subject of consciousness, and the nature of first-person content' (Peacocke 2012a: 109). According to these conceptions, the fact that a specific episode of consciousness occurs implies that there's a particular subject undergoing that episode. This much seems uncontroversial but it doesn't yet yield the desired support. How does the doubting subject knows that it's herself that's undergoing this conscious episode? Peacocke explains that 'only the subject whose thinking it is can be aware of the thinking in the distinctive way that stage (1) of the *Cogito* involves' (Peacocke 2012a: 112). And the reason for this is related to the specific nature of *de se* thinking. In virtue of undergoing such a specific episode the relevant thought necessarily refers to the thinking subject.

And any subject that employs the relevant concept of thinking in the right way has to be aware of that fact.

A.1.7 *The relation between authority and privileged access*

Not everyone agrees that the source of epistemic authority regarding self-knowledge originates in one's privileged access to one's own mind. Most prominent among the opponents of this connection are those who pin down authority in the linguistic convention of *avowals* (Brandl 2014; Wittgenstein 1953). We usually have to take the other's word when she's claiming to believe something. First-person utterances are a kind of speech act that has to be taken at face value, given there's no contradicting evidence.

A.1.8 *Self-deception*

The possibility of deceiving oneself proves problematic for many accounts of self-knowledge. If self-knowledge goes hand in hand with authority, then self-deceptive beliefs seem to undercut that relation. A subject who deceives herself into believing that p isn't authoritative with regard to her knowing that she believes that p. Bilgrami (2012) tries to deal with this challenge by distinguishing beliefs as dispositions from beliefs as commitments. The latter kind are important for the authority of self-knowledge. It's because subjects normatively commit themselves—by refraining from consciously holding or forming any incompatible beliefs—to certain beliefs that they're authoritative with respect to them. As a result, subjects with self-deceptive beliefs can really be said to hold these beliefs—thus explaining their authority and claim to self-knowledge. The supposed conflict between authority and self-knowledge on the one hand and the self-deceptive belief on the other is thus explained away. In the case of Beta's self-deceptive belief that she's charitable, she really does believe that she's charitable and she has authority over that item of knowledge. When outsiders then point to the uncharitable behaviour that doesn't fit her supposed charity we can explain her behaviour as being brought about by her uncharitable dispositions. But since these attitudes are of a different kind than her committed belief, there's no conflict with the authority in question. She really holds the belief that she's charitable. She just

has other mental states that bring about her uncharitable behaviour (cf. Coliva 2016: ch. 7).

However, this explaining away of the conflict between self-deception and authority isn't uncontested. Since claims to authority occur within the context of publicly communicated claims to knowledge, the exact status of these kinds of avowals is relevant. Wright (2001) distinguishes here between attitudinal and phenomenal avowals. The former concern claims about what a subject believes, desires, and so on. In contrast, the latter concern the phenomenological states that a subject is going through—such as her pain, what she's currently visually experiencing, and so on. Crucially, Wright argues that 'attitudinal avowals do not exhibit the strong authority of phenomenal avowals: to the extent that there is space for relevant forms of self-deception or confusion, sincerity-cum-understanding is no longer a guarantee of the truth of even basic self-ascriptions of intentional states' (Wright 2001: 324). In other words, the phenomenon of self-deception only has a home with regard to one's intentional attitudes. And in these cases it does in fact collide with the purported authority of self-knowledge.

With regard to the theory defended in this book, what should be preserved is a form of authority relative to the primitive self-ascriptions of a subject. These self-ascriptions are candidates for genuine self-knowledge because they have the same epistemic base as the phenomenal avowals that Wright distinguishes. This also has the consequence that there doesn't seem to be room for self-deception with regard to these self-ascriptions. Considering the possibility of being severely misguided with respect to one's actual possibilities of interaction with the world, this seems a strange claim. However, if a subject is misguided about the way she can interact with the world, this will be mirrored in her behaviour. She then might hold a false belief about the relation between her body and the world—as in cases of the rubber hand illusion or phantom limb—but not a deceptive one.

A.1.9 *Wittgenstein's attack on self-knowledge*

The condensed version of Wittgenstein's attack on self-knowledge of course doesn't do it real justice. However, since it's only marginally important to the general arguments of the book, I'll use this appendix entry to slightly expand on it and clarify some concepts. Wittgenstein's

main attack is on the idea that there's something like privileged access to our own mental states. From a naïve point of view, one could think that our words for phenomenal events—such as pain—are meant to describe and report these events. But Wittgenstein argues that this is a misguided way of describing it. He claims that 'the verbal expression of pain replaces crying and does not describe it' (Wittgenstein 1953: 89e). This is the *expressivist* idea that words don't state or report the sensations a subject undergoes but express them. There's no relation of representation between the word and the sensation in first-person avowals—rather the word is an expression of the sensation.

Wright (1998: 25) explains that 'it is the so-called private language argument (…) which targets the idea of phenomenal avowals as inner observation reports'. The main argument can be condensed as an observation of the fact that from the first-person perspective there's no distinction between what seems right to the subject and what is really right. If a subject examines her sensations in isolation, it doesn't make sense for her to think something along the lines of *It seems like I'm in pain but I'm not*. But if the expressions of our sensations are anything like regular observational reports, such a distinction between appearance and reality needs to be applicable. Hence, the verbal expression of seemingly private sensations doesn't amount to a report or assertion. This is supposed to undercut the claim that we have a better 'view' on our inner mental lives than others and that we're privileged because of that. Phenomenal avowals simply aren't assertions and thus don't amount to something that could qualify as knowledge.

The second point pertains more explicitly to claims of self-knowledge with regard to our intentional attitudes. Again, while it might seem that claims about what oneself believes or desires are mere reports made on the basis of introspection, this doesn't seem right. The main difficulty for such a view concerns 'the answerability of ascriptions of intentional states, like expectation, hope, and belief, to aspects of a subject's outward performance that may simply *not be available* at the time of avowal' (Wright 1998: 29). In other words, whether a subject can be properly said to believe that the train has arrived depends not only on the introspective world of the subject. It also depends on her behaviour because our beliefs are rationally connected with our intentional actions. As such then, the rationality of a belief expression can't merely depend on some kind of privileged epistemic position that supposedly

obtains. Consequently, such avowals aren't in any way epistemically distinguished—as the defender of a superior kind of self-knowledge would have us believe.

Neither of these attacks, however, troubles the account of *de se* thinking given in the book. It accepts the similarity between knowledge of the external world and knowledge of one's own mind. A subject's claim to knowledge about her own mental states is on par with other kinds of knowledge. Accordingly, a subject's self-ascription of the property of believing that p comes with the same epistemic requirements as the ascription of a property to some other object. One's own introspective access is only one more or less reliable source of knowledge of one's own mental states among others.

A.1.10 *More on reasons and the first person*

The claim that our action requires some *de se* attitude to the reasons that speak in favour of the action isn't universally accepted. Cappelen and Dever (2013) argue that agency doesn't require first-personal intentions and desires. Their main argument rests on the case of an omnipotent being which can bring about states of affairs just by intending them to be so-and-so: 'We think there could be a god, who can bring about states of the world just by intending them or maybe just by thinking them. The god thinks, "The door is closed," and straightaway the door is closed. On our view, this god's actions can be rationalized even if we don't specify any kind of *de se* state' (Cappelen and Dever 2013: 37). They instead propose what they call the Action Inventory model which takes subjects to have third-personal beliefs and intentions about states the world is or should be in. Furthermore, subjects have an 'action inventory' which they want to match with their intentions. Basically, if Alpha has the third-personal intention *Alpha has soy ice cream* and furthermore the action ⌜Alpha gets soy ice cream⌝ is in her inventory, then she will match these two and perform the relevant action. No *de se* belief or intention is supposedly required.

There are a number of problems with this argument. First, it's unclear how the case of the omnipotent god has any kind of relevance for the way intentional action is motivated and explained in normally potent subjects. Prosser (2015) argues along these lines that at least the typical actions of subjects and possibly all actions of normally potent

subjects require *de se* intentions. Hence, the fact that some omnipotent subject can bring about states of affairs without having first-personal intentions is irrelevant to our endeavour. Secondly, the Action Inventory model fails to explain why the intended states of affairs bear any motivational power and how they link up in the required way to the action inventory of a subject.

Furthermore, there are quite potent arguments to the result that subjects have to bring reasons for actions into some rational connection with their own attitudes. And that's to say that they have to take some kind of *de se* attitude towards the reason in question. Along these lines, Burge (1998) argues that 'to have reasons one must, I think, have had some tendency to have one's thoughts and attitudes be affected by them' (Burge 1998: 251). He holds that a subject can only have a reason—such as the fact that the plant is withering—to act in a certain way if the subject could be rationally affected by the reason's motivational force to do something in a certain way. And this is tantamount to standing in some kind of first-personal attitude towards the reason. The peculiar nature of *de se* thinking 'marks, makes explicit, the immediate rational relevance of invocation of reasons to rational application, or implementation, and motivation' (Burge 1998: 253). The rationality of action is connected to making subjects accountable for their reasoning and thus the *reasons* they invoke for their action. However, the impersonal way in which actions are connected with reasons in the Action Inventory model leave no room for these kinds of considerations. Thus, if we want to hold on to the possibility of actions being *rational*, they have to be accounted for in terms of *de se* thinking—even if not all subjects that act for reasons are responsive to questions of rationality.

A.2 ADDENDA TO CHAPTER 2: DIVIDE AND CONQUER

A.2.1 *When are two beliefs identical?*

It's not quite simple to clearly determine when two beliefs are identical. This is because that question can be understood in different ways. If we want to look at the specific tokens of beliefs—the particular mental states that a subject is in when she's believing something—then it seems impossible for two beliefs to really be identical. As such, it seems nonsensical to understand the question in this way. However, we can imagine science fiction scenarios where two subjects are hooked up in certain ways to share their mental lives. Here, we might still want to know whether they're thereby in the same particular mental state or whether they entertain two distinct particular mental states.

Normally, however, we take a different perspective on the question of belief individuation. We want to know under which conditions two particular belief tokens of two spatially or temporally distinct subjects are of the same *type*. So, we want to know whether my belief *I'm hungry* that I entertain on Saturday is the same as my belief *I'm hungry* that I entertain on Tuesday or your belief *I'm hungry* that you entertain on Friday. A definite answer to this question would need to delve into the different ways of characterising beliefs in general. Are they occurrent mental states? Can they be dispositions? Can they be both?

For the purposes of this chapter and the book the following more narrow focus is relevant: If beliefs are individuated in virtue of a proposition, then the proposition that's entertained in believing will be important to determine whether two beliefs are the same. So, if we think that Alpha's belief *I'm wounded* and her belief *Alpha is wounded* are both individuated by the proposition <Alpha, being wounded>, then her beliefs are identical. But if propositions care about more than these rock bottom conditions of satisfaction, we might say that the former belief corresponds to the proposition <the thinker of this thought, being wounded> and the latter corresponds to <Alpha, being wounded>. On such a view of propositions the two beliefs aren't identical because they're individuated by two distinct propositions.

There's a further way to look at how beliefs are individuated and thus how to answer the original question. We can take them to be characterised not just on the basis of these semantic aspects but also on

both epistemic and phenomenological aspects. Hence, we might ask: 'Does the subject draw the same inferences from this particular belief?' or 'Does the world phenomenally appear to the subject in the same way?' Both these questions severely widen the scope of the question.

So, the way we draw inferences might determine how beliefs are individuated (cf. Soldati 2016). Another aspect of the first question is illustrated by Frank Jackson (1999) who puts it into relation with the normativity of belief (Boghossian 2008; Brandom 1998):

> Someone who believes that P, and that if P then Q, *ought* to believe that Q. It is not simply that, by and large, they do believe that Q. It is that if they don't, there is something *wrong*.
>
> Jackson 1999: 421

Because beliefs have specific normative properties, the individuation of beliefs has to reflect that. Furthermore, we can also inquire about the phenomenology of our beliefs (Bayne and Montague 2011; Chudnoff 2015). Do some beliefs have a distinct kind of phenomenology? There's considerable disagreement about the question whether beliefs and other cognitive states have some kind of phenomenal aspect—and even if they had, whether that matters.

For our purposes the individuation of beliefs follows from a combination of all these aspects: semantic, epistemic, and phenomenological. More precisely, when Alpha's in a particular mental state M_1 that's phenomenally sufficiently similar to a particular mental state of Beta M_2, and the inferences they draw from these mental states are sufficiently similar, and the semantic content—understood in a context-dependent way—is the same, then M_1 and M_2 are the same. For instance, if Alpha believes *I'm going to miss the bus*, this belief is characterised by three aspects. The content of the belief is something like <the thinker of this thought, going to miss the bus>. The epistemic nature of the belief is connected to her then intending to run or calling someone that she's going to be late, maybe feeling ashamed, and so on. And finally, the phenomenology might be related to some feeling of dread or fear. Beta's belief *I'm going to miss the bus* can be said to be the same if these three aspects are sufficiently similar.

A.2.2 *Fregean and Russellian propositions*

The two mentioned different ways of characterising propositions corresponds to a distinction between *Fregean* and *Russellian* propositions. The latter—also called *singular* propositions—have as their constituents the objects and properties themselves, whereas the former are constituted by the senses in which these objects are presented to us in thinking.

One argument for the necessity of Russellian propositions comes from the discussion about indexicals and demonstratives. Utterances and mental states which involve these terms are supposedly hard to capture with the Fregean notion of a sense. We'll see some of these arguments resound in the main text. Kaplan (1989) and Perry (1977) argue accordingly that they require a treatment which involves singular propositions. On the other hand, there are philosophers defending the Fregean approach against these kinds of attacks (cf. Chalmers 2011; Evans 1985; McDowell 1984). For our purposes, nothing important hinges on this dispute because the question of the right model of proposition is secondary to the more important epistemic question. We don't have to decide the metaphysical and semantic question about the right kind of proposition before we have clearly understood what exactly is involved in the nature of *de se* thinking. Once this is done, some theories of propositions might prove more or less fruitful than others.

A.2.3 *The relation between the first and second element of belief*

There are different views about how the intentional object—i.e. the real existing thing we're thinking about—is related to the way we're thinking about that object. In the case of seeing Wonder Woman, it seems that the intentional object itself, our superheroine, is presented to us in our mental state in a certain way when we see her as the wielder of the Lasso of Truth. In such a case, it makes sense to say that the second element of belief is how the first element is presented to us in thinking. It's Wonder Woman herself that's presented to us as the wielder of the Lasso of Truth.

However, in other cases, we might want to refrain from saying that the intentional object itself is presented to us in thinking under a cer-

tain mode of presentation. So, in believing that the wielder of the Lasso of Truth is smart, we might say that the intentional object of my belief is a certain Russellian singular proposition. This might tempt us to say that my thought is about the proposition which involves the particular lasso wielding person who instantiates the property of being smart. However, the way I'm thinking about Wonder Woman in that belief shouldn't be characterised as thinking about a certain proposition in a certain way. Rather, we want to say that I think about a certain woman in a certain way.

Hence, it's not generally true that the second element of thinking is characterised by the way we think about the first element of thinking. Sometimes, we need to say that it's characterised by the way we think about certain constituents or parts of the first element. The mode in which Wonder Woman is presented to us in believing *The wielder of the Lasso of Truth is smart* determines the second element of thinking and not the way that proposition is presented to us.

Furthermore, describing the relation in this way doesn't imply a specific semantic or epistemic directionality. While it might seem that we 'start' from the second element of belief and then converge on the truthmaker of that belief—our first element—this isn't necessarily the case. The only thing that's implied is that we can differentiate these different elements or aspects of a belief. A related question is then whether we have modes of presentation 'in mind' or the actual objects. With respect to perception, this debate is often lead under the labels of *representationalism* (Siegel 2011; Tye 1995) and *direct realism* (McDowell 1996; Soldati 2012).

A.2.4 *Contextualism and context*

The examples given in the text might imply that there are two distinct kinds of belief which can be clearly distinguished. One being independent of context and another dependent. Furthermore, you might think that only *de se* beliefs and others which involve indexicals like 'here', 'now', or 'this' are in the latter category. This doesn't follow and is highly contentious. In fact, both metaphysical and semantic contextualists might hold that all kinds of beliefs are highly dependent on their context.

The exact extent of contextual influence on our beliefs is a highly contentious and debated matter which can't and needn't be settled here. However, it's important to clarify some claims that can be made in this area. On the one hand, there's the idea that the truth of an utterance or mental state is dependent on the context. According to this idea, one and the same type of utterance can be true in one case and false in another. Here we find the source of relativist claims about truth and morality. On the other hand, one can bracket the normative question and merely operate on the level of individuation of beliefs and utterances. We might say something like 'Some utterances can only be understood if we know their context'. For instance, a subject can only understand what's meant by someone saying 'She's here' when she knows who's referred to by 'she' and which place is picked out by 'here'. Of course, this latter claim can be related to a claim about the relativism of truth.

A further issue concerns the exact definition of a 'context'. Aren't possible worlds just special cases of contexts? Couldn't we also say that contexts need to be included in the characterisation of what a possible world is? Here, I propose to draw a distinction between possible worlds and contexts on the line of authors like Kaplan (1989), Lewis (1998), and Stalnaker (1999). Accordingly, contexts are characterised by parameters like place, time, and subject. On the other hand, possible worlds are to a degree insensitive to these differences and sum them up. We find different contexts in one and the same possible world because a possible world can encompass many distinct place–time–subject-triples.

A.2.5 *More on the paradox of self-consciousness*

It's important to slightly qualify this argument in two ways. First, does the acquisition of the first-person pronoun in fact presuppose *de se* thinking? And secondly, is Bermúdez' paradox actually intended against the Kaplanian theory at all?

One could argue that a subject can learn the meaning rule of the first-person pronoun without knowing that her own application refers to herself. In other words, one might try to show that the acquisition of the first-person pronoun isn't dependent on a prior capacity for *de se* thinking. The problem with such an attempt to salvage the linguistic

approach is that it doesn't explain how subjects come to think in the *de se* way. Imagine a subject who knows the rule '"I" always refers to the speaker' but doesn't know that her own use of 'I' refers to herself. If such a subject is possible, it wouldn't fall under the scope of the linguistic approach which only cares about subjects thinking in the *de se* way. Or, to put it into the words of Anscombe: 'The explanation of the word "I" as "the word which each of us uses to speak of himself" is hardly an explanation!—At least, it is no explanation if that reflexive has in turn to be explained in terms of "I"; and if it is the ordinary reflexive then we are back at square one' (Anscombe 1975: 48).

Furthermore, one might take this as undermining the paradox because it shows that knowledge of the character of 'I' doesn't presuppose the capacity for *de se* thinking. Unfortunately, the knowledge of the character is only insofar relevant for *de se* thinking as it explains a subject's use of the first-person pronoun to express her *de se* attitudes. And such an expression requires the capacity to think in the *de se* way because it requires that the subject is aware that *she herself* is the producer of that sound or the speaker of that utterance.

This should further illuminate how we should react to the second question. While Bermúdez doesn't construe his argument explicitly as a reaction to Kaplan's theory, it's a direct response to what he calls the 'deflationary theory' which is mainly characterised by the claim that 'once we have an account of the semantics of the first-person pronoun, we will have explained everything distinctive about the capacity to think thoughts that are immune to error through misidentification' (Bermúdez 1998: 11). This characterisation of the deflationary theory is reasonably similar to how we characterised the linguistic approach to *de se* thinking. As such, it's reasonable to take Bermúdez' paradox as a suitable response to our target.

A.2.6 *Frege on concepts*

The way that Frege uses the term 'concept' is quite technical and not always helpful. Coming from his treatment of logic—which he intends to rid of all reference to psychological states—he introduces the term 'concept' as a purely 'objective' notion and wants to distinguish it sharply from the subjective associations that are present in the mind of the subject. In this vein, he writes in his *The Foundations of Arithmetic*

(1953) about the relation between subjective idea, objective content, and the object of thought:

> An idea [*Vorstellung*] in the subjective sense is what is governed by the psychological laws of association; it is of a sensible, pictorial character. An idea in the objective sense belongs to logic and is in principle non-sensible, although the word which means an objective idea is often accompanied by a subjective idea, which nevertheless is not its meaning. Subjective ideas are often demonstrably different in different men, objective ideas are the same for all. Objective ideas can be divided into objects and concepts.
>
> Frege 1953: 37, fn 1

So, Frege presents us with the following theory. Every intentional attitude involves a subjective idea, an objective concept and an object. The part of this attitude that is communicable in language is fully exhausted by the concept and the object. These are the things which aren't different from one subject to another and thus constitute the meaning of a word. For instance, if a subject forms the judgement *There is a horse*, she employs the concept of a horse—determined by the conditions under which an object is the referent of that concept. And she communicates this by using the word 'horse' with its perfectly objective meaning. The specific way she pictures the horse in her mind isn't part of the concept and the meaning.

This also sheds some light on how to understand Frege's famous remark on the first person which doesn't really square well with the idea that sense determines reference: 'everyone is presented to himself in a particular and primitive way, in which he is presented to no-one else' (Frege 1956: 298). If the sense of the first person concept was guided by this primitive way of being presented to oneself, then it couldn't play the objective role it's supposed to play. After all, the primitive way is supposedly private and not communicable. Instead, we have to distinguish the subjective idea—the primitive way we're presented to ourselves in thinking—from the sense of the first person concept.

It's only the objective first person concept that's relevant in determining the content of an intentional *de se* attitude. And this concept, in turn, is fully exhausted by the sense which determines the rules of reference: Whenever it's applied, it refers to the thinking subject.

However, this still allows that a subject comes to know that she herself is the thinking subject on the basis of her subjective, particular, and primitive idea of herself. So, the subjective idea is a way of epistemically connecting one's application of the first person concept with one's grasp of oneself in the *de se* way.

This clearcut distinction between the 'subjective' phenomenal state and the 'objective' concept which is communicated isn't without critics. For instance, Michael Dummett speaks of the 'false dichotomy between mental images as subjective and incommunicable, sense as objective and communicable' (Dummett 1973: 158). Another external critique comes from the empiricist and representationalist traditions. These argue that concepts have to be concrete things—as opposed to abstract Fregean senses—in order to play certain roles that concepts need to play in thinking (Margolis and Laurence 2007). Neither propositions nor concepts are independent eternal entities which subjects 'latch' on to in their thinking. Rather, these things are first and foremost constituted through the intentional states that subjects are in. This 'intentionality first' view of propositions and concepts is also often associated with Frege's most dire antagonist Husserl (1983, 2001).

To illustrate the conceptual approach to *de se* thinking, we won't take Frege's theory as a template but rather focus on strategies that have been inspired by this focus on concepts and their fundamental rules of reference.

A.2.7 *One or many first person concepts?*

Why should we think that there are individual first person concepts which are distinct from person to person? One reason is that the public concept, which is individuated by the reference rule 'the referent of "I" is the subject of the thought containing the application of the first person concept' (cf. Burge 1998: 246), can't guide the application of the concept. As we saw, the understanding of this rule presupposes some form of *de se* thinking. Frege wanted to do justice to this problem by introducing the 'particular and primitive way' (Frege 1956: 298) in which everyone is presented to herself.

This has lead some people to the claim that we need to have something like an individual first person concept to account for this initial grasp of the concept. Along these lines, Kapitan (2016) argues:

'the *I* concept is instrumental in the initial self-identification and the apprehension, not something employed subsequent to a first-person thought. It is primitive in the sense of not containing another singular sense as a component' (Kapitan 2016: 315). So, it's wrong to say that the first person concept is applied only once the subject employs the reference rule. If we take the first person concept as described by the reference rule, we're clearly confronted with a complex concept which isn't primitive in the required sense. It's rather the other way round: We need an individual mode of presentation—a private sense—that guides the application of the public concept. Of course, Frege was extremely reluctant to accept such a private sense.

Another route to individual first person concepts was established by Hector-Neri Castañeda, who argued for the now widely accepted view that first-person thinking is irreducible to some kind of descriptive thinking (Castañeda 1999a,c). Most strikingly, in discussing Chisholm's view that there are individual essences, so-called *haecceities*, which are associated with the way subjects are presented to themselves in *de se* thinking, Castañeda argues that 'first-person individual concepts seem to be private to each person, and cannot be thought by others in the way Chisholm envisioned' (Castañeda 1999b: 120). So, contrary to what Chisholm (1976) thought, we can't at the same time hold on to the idea that there are individual concepts and that all concepts are necessarily intersubjective. One or the other has to go.

A.2.8 *Mental states and functionalism*

Functional approaches have quite the tradition in the philosophy of science and the philosophy of mind of the 20th century and I can't do that tradition full justice here. Much of the tradition goes back to Ramsey (1931) who has inspired so-called *Ramsey sentences* which are existentially quantified sentences that describe only the relations between the related terms. These sentences are intended to distinguish clearly between observable and non-observable things (cf. Carnap 1950). It's not so clear whether this attempt was successful, with its most pronounced critic being Quine (1953).

Within the question of the individuation of mental states, the rise of computing machines has given additional drive to the quest of functionalism. The general idea is that a mental state isn't characterised by

its internal structure—such as the proposition that's believed or desired or the non-observable phenomenology in question—but on the basis of its functional role. Two cognitive states are identical just in case they have the same functional profile. And two states have the same functional profile if they produce the same output with the same input. Lewis (1966) has aptly characterised this general strategy:

> The identity theory says that experience-ascriptions have the same reference as certain neural-state-ascriptions: both alike refer to the neural states which are experiences. It does not say that these ascriptions have the same sense. They do not; experience-ascriptions refer to a state by specifying the causal role that belongs to it accidentally, in virtue of causal laws, whereas neural-state-ascriptions refer to a state by describing it in detail. Therefore the identity theory does not imply that whatever is true of experiences as such is likewise true of neural states as such, nor conversely.
>
> Lewis 1966: 19

A.2.9 *Do we need direct importance?*

One might argue that the characterisation of the functional approach against Perry's background is slightly misguided: Not all instances of *de se* thinking are of 'direct importance' to the thinking subject. How is this objection supported? We've already seen that Perry develops the concept of a self-informative way of knowing (A.1.5). A given mental state S is a self-informative way of knowing for the subject that she is ϕ if being in that state 'normally carries the information that the person in state S is ϕ' (Perry 2002: 204). Certainly, information that's gathered through this channel is normally about the subject—if we bracket science fiction scenarios—and thus usually of direct importance for the subject. But, so the argument continues, not all forms of *de se* thinking coincide with this way of knowing. Some *de se* intentional attitudes are formed on a different epistemic basis. Remember that Perry needs self-informative ways of knowing to account for immunity and not for all kinds of *de se* thinking. And since not all instances of *de se* thinking are immune to error through misidentification, not all *de se*

beliefs are of direct importance to the subject. Hence, the functional approach is ill-defined in its reliance on direct importance.

This argument against the chosen characterisation of the functional approach fails for two reasons. First, the fact that not all instances of *de se* thinking are based on a self-informative way of knowing doesn't imply that other *de se* attitudes aren't of direct importance to the subject. Even if a subject forms her belief *My legs are crossed* on an epistemic basis which doesn't come with immunity, her belief is nevertheless of direct importance to her. For instance, she takes that as an immediate and direct reason to uncross her legs before standing up. And this is exactly the functional role of *de se* attitudes.

Secondly, conceding that only some *de se* attitudes are formed on the basis of these first person methods of knowing doesn't yet imply a general functional difference between those *de se* attitudes which are formed thusly and those which aren't. We shouldn't be fooled into saying that only the former kind are equipped with direct importance for our behaviour. For instance, Récanati (2012) builds his theory of *mental files* on what he calls 'epistemically rewarding relations'—inspired by Perry's framework. The relevant *self* file is supposedly formed on the basis of something like the self-informative way of knowing. It gets all its characteristics from this specific epistemically rewarding relation. And this includes the feature of being of direct importance to the thinking subject. However, once the file is established, other kinds of information—which might be gained in a different way—can also be stored there. The fact that the information is stored in that particular mental file equips it with that particular kind of cognitive function. As Récanati writes: 'There is much information about myself that I cannot get in the first person way (…). That information goes into my *self* file, however, because I take it to concern the same person about whom I also have direct first-person information, namely myself' (Récanati 2012: 36, fn 8). My belief *My legs are crossed* is of direct importance to me because I therein employ the self file. And this file was created to store information that's supposed to be about me.

A.3 ADDENDA TO CHAPTER 3: BACK TO THE PRIMITIVE

A.3.1 *What's an ascription of a property?*

There's a certain unease concerning the way I described a newborn as ascribing a property to the stick. In particular, it concerns the question whether we really want to allow non-conceptual thought to be described in terms of ascription of properties. Several questions are of relevance here. First, does the newborn really ascribe a property to an object in seeing the stick as being bent? Secondly, what's the relation between the ascription of a property and the instantiation of a property? Thirdly, is there any epistemic import in the ascription of a property? How does it relate to knowledge? And finally, what does it mean to say that ascribing a property to oneself amounts to 'placing oneself into a group' of objects?

I'm not going to give a complete answer to all these questions since I hope that it's possible to remain somewhat neutral on some of them. Nonetheless, the nature of property ascription needs to be made a bit more concrete. There are two worries looming in the background which bring about the four questions above. The first worry is that non-conceptual thought isn't properly described as having an object–property structure. Accordingly, we shouldn't describe the newborn as ascribing a property *to an object*. The second worry concerns the possibility of thinking in terms of properties at all without conceptual abilities. In other words, we might want to say that a subject ascribes a property just in case she uses the relevant concept. Since the newborn doesn't have the concept of a bend, it can't ascribe the property of being bent to the stick.

The reason why it's difficult to give clearcut answers to the questions above is that the notion of non-conceptual content—from which much of the foundation for the worries derives—is itself not clearly defined. For instance, Levine (2016) explains that 'for an experience to qualify as having nonconceptual content, it must have *representational* content, content in which features of the *world* are represented by an experiencing subject' (Levine 2016: 856). This suggests that property ascription—understood as potentially non-conceptual—involves the representation of the world involving objects. On the other hand, we might characterise property ascription as 'structure-implicit, which is

to say that it contains no explicit representation of subject/objects and predicates' (Musholt 2013: 653). This presents us with the opposing idea: there's no object involved in the ascription of properties.

Against this background of disagreement, I propose the following answers to the questions. First, we can describe the newborn as ascribing a property to the bent stick insofar as she interacts with the stick *as if* it were bent. In other words, saying that the newborn ascribes a property to an object is a way of describing the situation in functional terms. The newborn would interact with the stick differently if she saw it as being straight. And this would amount to her ascribing the property of being straight to the object. This way of putting it remains neutral regarding the question whether the infant thinks about the world in terms of objects and predicates, in terms of affordances (Pettit 2003), or in terms of placing a feature (Strawson 1959).

Secondly, ascription of a property resides in the realm of appearance since a subject can behave *as if* a property were instantiated even if that property isn't actually instantiated. Saying that a subject ascribes a property to an object doesn't commit us to the claim that the property is really instantiated or can be instantiated in that object at all. Hence, in a hallucination, I might ascribe the property of flying to an elephant despite the fact that elephants can't fly in the actual world. In this, then, ascriptions of properties are similar to the applications of concepts. In both cases, the cognitive state is prone to misrepresent how things really are. However, since we can reasonably say that newborns ascribe a property—in the sense elaborated above—without applying the concept, we don't have to claim that the possibility of property ascription presupposes the possession or application of concepts.

Thirdly, ascribing a property to an object is something that can be put into rational connection with other mental states. This means that ascribing the property of being poisonous to the spider warrants the ascription of the property of being dangerous to that same spider because poisonous things are usually dangerous. This epistemic reading of property ascription also applies to the weak implications of the functional 'redescription'. Interacting with an object *as if* it were poisonous is a good reason to interact with that thing *as if* it were dangerous. Hence, the ascription of properties can be subject to rational and epistemic norms and give rise to knowledge in those cases where it's viable to speak of a subject as *knowing* something.

Fourthly, since properties are logically understood as sets of objects, I sometimes speak as if subjects believe the apple to belong to the group of tasty things or that the subject thinks that the apple instantiates the property of being tasty. Of course, this is an overintellectualisation *par excellence*. Subjects don't need to entertain these complex intentional attitudes in order to ascribe a property to an object. Rather, ascribing the property of being tasty to the apple is logically equivalent to 'placing the apple into the set of tasty things' or 'thinking of the apple as instantiating the property of being tasty'. However, these logical translations don't correspond to actual cognitive translations.

A.3.2 *Lewis on properties*

These kinds of properties might strike you as quite weird and rather far-fetched. However, it shouldn't surprise you that Lewis didn't care about the fact that his properties are outlandish in that way. In his most famous piece of philosophy, *On the Plurality of Worlds* (1986), he writes that 'the abundant properties may be as extrinsic, as gruesomely gerrymandered, as miscellaneously disjunctive, as you please. They pay no heed to the qualitative joints, but carve things up every which way' (Lewis 1986: 56). Nonetheless, Lewis is well aware of the fact that some properties are more firmly rooted in the world of thinking and causality. He therefore introduces a distinction between *sparse* and *abundant* properties (Lewis 1983a). While the abundant properties—as the quote explains—are running wild, sparse properties are those which are needed to account for the similarity and causal relations between various objects. For instance, *being coloured* is a sparse property while *inhabiting the actual world* might not be.

Lewis is usually taken to be a proponent of *nominalism* about properties. This means that properties don't correspond to actual mind-independent entities, as realists would claim. Properties aren't real in the sense of corresponding to individual independent things. Among the more famous realists about properties are Plato and Aristotle. While they disagree about the exact nature of properties—most importantly about the question whether properties can exist uninstantiated—they concur that properties are concrete individuals. In contrast, Lewis reduces properties to sets of all possible instances. Hence, properties don't exist as independent entities but are agglomerates of other things.

Because this way of understanding properties is intimately tied with Lewis's realism about possible worlds, it inherits many of the problems which vex this unusual view (Egan 2004).

Finally, we also have to distinguish Lewis's view on properties from *conceptualism* about properties (cf. Cocchiarella 2007). While conceptualists also hold that properties aren't mind-independent, they argue that the expressions we use for properties—such as 'dampness' or 'strength'—actually just refer to the concepts we use in thinking. Accordingly, properties are in a sense reduced to concepts. It's clear that Lewis doesn't hold such a view because concepts aren't sets of possible individuals and are therefore distinct from Lewis's properties.

Does the theory defended in this chapter depend on any particular view of properties—most importantly Lewis's? I don't think so. Two reasons speak for a less committal theory of properties working in the background. The first is that I don't have to buy the story about abundant properties. Since the lived body account doesn't attempt a reduction of all thinking to *de se* thinking, we can rest content that only sparse properties are self-ascribed. Moreover, it's even possible to refer back to propositions for other theoretical jobs. The only important claim is that the ability of subjects to think in the *de se* way has to be understood in terms of self-ascribed properties and not entertained propositions. The second reason is that self-ascription of properties can be understood using quite different metaphysical theories of properties—be they nominalist or realist in nature. As long as it makes sense to speak of subjects as ascribing properties in thinking, we can accept different plausible and compatible accounts of the the nature of properties.

A.3.3 *Additional criticism of the property theory*

Given the overall dominance of propositional theories, it's no surprise that the property theory hasn't been embraced with open arms by the philosophical community. There are a number of reasons why this is the case. Most of them derive from the property theory's attempt to unify the picture of intentional attitudes such that every belief corresponds to the self-ascription of a property. However, it's doubtful whether the property theory is up to this task. Let me just illustrate two problems that originate from this and discuss some repercussions

for my defended lived body account. The first problem concerns the relation between *de se* beliefs and communication. The second is about the fact that our beliefs are usually true or false.

Regarding the first problem, Robert Stalnaker (1999) presents us with the following dilemma for the property theorist. Either the objects of assertions are self-ascribed properties or they're propositions. In the former case, people seem to talk about themselves when they're asserting something. This is bad, because some of our assertions are about other things and thus their truth-conditions don't concern us. In the latter case, the property theorist needs to bite the bullet and claim that assertions aren't a direct expression of our intentional attitudes. Let's look at the relevant passage directly:

> If assertions are always self-ascriptions of properties, then people talk only about themselves. Alternatively, Lewis might hold that speech acts, unlike attitudes, have propositions rather than properties as objects. But then he must deny that speech is a straightforward expression of thought—that what a person says, when she believes what she says, is what she believes.
>
> Stalnaker 1999: 147

The second problem is that beliefs and other intentional attitudes are usually either true or false. Since, according to the property theory, the content of beliefs is a specific property and not a proposition, it's difficult to see how it can account for this semantic feature. If beliefs are relations to propositions, then they can be said to merely inherit the truth-value of the relevant proposition. But since properties—contrary to propositions—aren't true or false, the property theory can't claim that beliefs inherit their truth-value in the same way from their content. Neil Feit, a contemporary proponent of the property theory, attempts a rebuttal of this objection by claiming that 'talk about truth is appropriate for properties, insofar as we speak of properties as being true of their instances, e.g., *being clever* is true of every individual who is clever, and false of every one who is not' (Feit 2008: 16).

Independently of the question whether these two problems and their respective solutions are convincing, the lived body account is somewhat untouched by either. Since it doesn't entail a necessary reduction of all intentional states to the *de se*, it can accept that propositions have

some role to play in the way that subjects interact with their worlds and the people living in it. As such, communication might just imply something like a proposition that's expressed when we utter *de se* sentences. The bullet that needs to be bitten might not prove very hard. It's not necessary that our assertions *directly* express our mental states.

Additionally, we don't need to make up a story of how properties are in some sense true or false. We might just admit that there's no sense of talking about truth in the case of properties. At the same time, the truth or falsity of beliefs is easily accounted for in other terms. Alpha's belief *I am F* is true just in case Alpha has the property *F* or just in case the proposition <Alpha, being F> is true. There's no need to ban propositions from our picture.

A.3.4 *Centred worlds*

It's sometimes argued that this analysis is too crude and doesn't take into account the full-fledged Lewisian theory. The standard interpretation of Lewis's theory is that we self-ascribe what are called *centred worlds* whenever we're entertaining *de se* beliefs. However, it's unclear which kind of centred world Lewis actually subscribes to. The orthodoxy on this matter is the idea that centred worlds are pairs of a possible world and a centre. It's quite obvious that this explanation is flawed because it neither tells us what a possible world is nor what a centre is. The important point here concerns the question: 'What is a centre?' Again, the standard answer is that the centre is a pair consisting of an individual and a time (Liao 2012). So, when Alpha believes, on October, 31, 2016, that she's happy, the centre of her centred world is the pair that consists of Alpha and the day October, 31, 2016.

Now, it's obvious that this doesn't give us a proper *de se* belief. This is because Alpha can believe that she's happy without knowing anything about the date of her belief and while being oblivious to her own name and appearance. There's thus a need to identify oneself with a certain centred world and not another. An answer which is plausibly supported by the few things Lewis has to say is that this identification is based on the notion of identity (cf. Liao 2012: 313). In fact, much of the discussion about the role of centred worlds assumes that we can simply treat them as the standard objects or contents of beliefs. This is too simple. Properly understood, a centred world designates a certain

individual with a property. For instance, Alpha believing *I'm happy* amounts to her ruling out all the possible worlds where the centre doesn't have the property of being happy. However, she still needs to self-ascribe that centred world if it should serve as a basis for *de se* thinking. Or, as Richard Holton (2015) argues:

> The first thing to note is that, as Lewis presents things, the role for centered worlds is to stand in for properties. So we still need the idea that they are self-ascribed. That is something that is obscured in much of the subsequent literature, where centered worlds are often taken to play the role of propositions, that is, as things that are straightforward objects of belief, rather than of self-ascription. Introducing centered worlds does not change the fundamentals of the account; it just changes the way we describe it.
>
> Holton 2015: 403

In other words, for a centred world to play any role in *de se* thinking, it has to be self-ascribed. Alpha might have identified all the worlds where the centre has the property of being happy, but she still needs to take herself as being one of these individuals. And that's tantamount to self-ascribing a centred world. Hence, a more technical redescription of the Lewisian theory doesn't change the general layout of the property theory which is that subjects self-ascribe properties when they think in the *de se* way.

A.3.5 *Self-ascription under identity*

You might argue that this reconstruction of Lewis's idea isn't quite charitable. Let me try to convince you that it's the only way we might read the property theory such that it can explain the peculiarity of *de se* thinking (Wüstholz 2018). I argued that *de se* thinking amounts to an ascription of a property under the relation of identity while *de re* thinking amounts to an ascription of a property under some other acquaintance relation. This much can be supported by looking at how Chisholm (1981) distinguishes between direct and indirect attribution.

If we now compare the two beliefs *I'm happy* and *Beta is happy*, we see that the content of these two beliefs is different. In the *de se* case, a subject ascribes the property of *being happy* while in the latter, she ascribes

the property of *being such that Beta is happy*. But this isn't enough to give us a clear distinction because the content alone makes both of these beliefs of the same type. They both look like *de se* beliefs because they both amount to self-ascriptions. But this is clearly not what we want. Furthermore, one and the same property could be the content of a *de se* belief in one case and the content of a *de re* belief in the other. Thus, the content alone isn't sufficient for our purposes and we have to move away from the content to the way that content is grasped—or how the property is ascribed.

The suggestion is then that the relevant acquaintance relations are capable of drawing the required distinction. Beliefs *de se* are ascriptions of properties under the relation of identity while beliefs *de re* are ascriptions of properties under some other relation of acquaintance. Now, you might argue that the subject doesn't have to think of herself *as* the thing identical to herself in self-ascription. The metaphysical relation of identity can be exploited in her self-ascription and doesn't have any epistemic import. So, the subject doesn't need to judge *I am such that the person identical to me is F* but simply *I am F* while exploiting the fact that she's identical to herself.

The problem with this attempted rescue is that the mere fact of identity can't be explanatory. For that fact obtains whether the subject self-ascribes in the *de se* or in the *de re* way. Hence, we require an explanation of how the exploitation of that fact can make some self-ascriptions cases of *de se* belief. The only plausible way of spelling this out, which results in *de se* thinking having the epistemic qualities it has, is that subjects have to take an epistemic stance on the fact that they're identical to themselves. Hence, the subject has to ascribe a property to an intentional object *as* the thing identical to herself. And this way of explaining *de se* thinking and self-ascription obviously doesn't work because it introduces the well-known regress problem.

Furthermore, it would be a misunderstanding of the argument if you took the claim to be that subjects really do think of themselves under the epistemic relation of identity in *de se* thinking. They don't; and that's exactly the point. The only way of making sense of the property theory under the banner of Lewis and Chisholm is to assume some epistemic relation underlying *de se* thinking. But the problems with this way of describing *de se* thinking shows that the property theory as such can't be accurate.

A.3.6 *The infinite regress in intentional action*

This regress is described in a similar way by Simon Prosser, who argues that what he calls 'first-person redundant' mental representations are essential for intentional action. These mental representations are nothing but irreducibly *de se* mental states. He argues that, in order for a subject to act on a specific thing in the world, the subject has to be aware of her relation to that object. Most importantly, the subject has to be aware of the kinds of relations between herself and the object, which determine possible interactions with that thing. Prosser calls these epistemic relations that are determined by the possibilities of interaction 'subject-environment' or 's-e relations'. Without a subject's awareness of these relations, she wouldn't know which actions she can perform with respect to a specific object in her surrounding.

However, merely knowing that Alpha stands in a particular relation to the glass isn't enough for Alpha to know how she herself can behave in order to interact with that glass. Prosser explains this point in more detail: 'Knowing how to act is not the same as knowing how *S* can act, even if I am in fact *S*. Ordinary knowledge that *S* stands in a s-e relation *R* to *o* gives me the right information, but it does not give it to me in the right form. Instead, for information about my s-e relations to enable me to act, their representation must be first-person redundant.' (Prosser 2015: 226)

He then proceeds to give a regress argument for this claim. Basically, Alpha can represent the fact that Alpha stands in a certain relation to the glass either in a first-person redundant—i.e. irreducibly *de se*—or in a non-redundant way. In the former case, she would express her representation as 'the glass is to the left' while in the latter case, she would express it as 'the glass is to Alpha's left'. Crucially, this latter representation wouldn't move Alpha to act on the glass without Alpha representing the fact that she stands in some relation to Alpha—for instance by being identical to her.

Now, this new representation can again be either first-person redundant or non-redundant. In the latter case, Alpha 'would be in the same representational state, and the same epistemic situation, as anyone else who wanted to make' (Prosser 2015: 227) Alpha act upon the fact that the glass is to Alpha's left. Hence, if we don't break the cycle through accepting an irreducibly *de se* state, we would require a new

representation of the relation that Alpha stands in with regard to this new fact. He therefore concludes that action is only possible on the basis of a first-person redundant representation.

A.3.7 *Holton on primitive self-ascription*

The argument for primitive self-ascription I developed raises some questions. A first one is exegetical and asks whether Lewis and Chisholm aren't themselves operating with a primitive notion of self-ascription. A second one concerns the claim that self-ascription does the necessary epistemic work within the property theory and not the more fancy, controversial, and attractive claim that the contents of mental states are properties necessary to account for. And finally, we might want additional support for the claim that self-ascription has to be understood as an epistemically primitive relation in which the subject takes herself to have certain properties.

It's quite difficult to answer the question whether Lewis thought of self-ascription as primitive or not. The standard reading is in terms of centred worlds which doesn't make any reference to either self-ascription or the specific epistemic nature of *de se* thinking. However, Liao argues that the standard Lewisian account was probably not Lewis's account: 'Considering Lewis's statements elsewhere and his other theoretical commitments, it seems that he in fact endorses the primitive identification account, and not the Lewisian account' (Liao 2012: 295). According to this reading, we can take Lewis as accepting the centred world reading of properties while at the same time requiring some epistemically primitive identification with a particular centred world over another. This might be tantamount to the primitive self-ascription account I've argued for, but, considering the rather big interpretational diversity, it's difficult to be sure.

The question of the epistemic role of self-ascription finds additional support from Cappelen and Dever (2013: ch. 5). While the authors—along with other *de se* skeptics such as Magidor (2015)—are rather skeptical about the essentiality of *de se* thinking, they argue that the central insight of Lewis was the fact that subjects have to *self-ascribe* properties in *de se* thinking. So, despite the focus on the claim that properties have to supplant propositions as the content of intentional

attitudes, it's actually the specific nature of self-ascription and its inherent epistemic primitiveness that explains how *de se* thinking is special.

Finally, additional arguments for the primitiveness of self-ascription are found in the already mentioned paper by Liao (2012) and also in Holton (2015), who argues that only primitive self-ascription can do the work required in Lewis's theory:

> As I have already stressed, taking self-ascription as primitive is crucial to Lewis's account. We normally think of ascription as a two-place relation: one ascribes a property to a thing. Self-ascription would then be the special case where the thing is the self. But that won't do the work here. If the self is just thought of extensionally, then we would have no way to distinguish the belief that one's pants were on fire from the belief that the pants of someone, who is you though you don't realize it, are on fire. (…) So we have to think of self-ascription as a one-place relation: one simply self-ascribes a property.
>
> Holton 2015: 403

Self-ascription has to be understood in a primitive way because any other epistemic relation would require an additional epistemic basis that justifies the subject in her self-belief. This is because a subject has to be warranted in believing that she's thinking about herself. However, every extensional way of individuating the subject results in the possibility of doubt concerning the question whether I'm the intentional object of my thinking. But, this kind of doubt isn't appropriate in the *de se* case. Hence, self-ascription—understood as the basis of *de se* thinking—has to be epistemically primitive.

A.3.8 *On the reductionist future*

There are a couple of things you could additionally say, if it's important to you. For instance, you could say that thinking *de dicto* is a case of thinking *de se* in the sense that it requires some sense of self-location in the logical realm. A subject thinking that Beta is tall has to understand in some sense that the fact that Beta is tall applies to herself when she believes it. The proposition <Beta, being tall> isn't just entertained as a mere possibility among others, but it is *believed*. Hence, the subject

could be said to take herself to inhabit a world where that proposition is true. So, you might want to claim that thinking *de se* is a prerequisite for thinking about other things in the world because every belief self-locates the subject in logical space.

While I sympathise with the general idea that thinking in the *de dicto* way is somehow related to thinking in the *de se* way, I don't have a clear way of spelling out the dependency between one and the other. There is certainly a primacy of the *de se*, but this doesn't yet imply that thinking *de dicto* is nothing but thinking *de se* in a special garment. The dependence might be conceptual, it might be epistemic, it might be logical, or it might be biological and therefore contingent.

A.4 ADDENDA TO CHAPTER 4: ORIGINS

A.4.1 *Wittgenstein and the nature of the subject*

What parts of Wittgenstein's view on the nature of subjects are adopted for the defended view? The important distinction that we're pointed to in the analogy to the eye is the difference in thinking about a subject *as a limit* and thinking about a subject *as an object*. Thinking of something as an object implies that one grasps it as belonging to a world that includes other objects with which it stands in certain relations. As Merleau-Ponty writes, thinking of something as an object requires apprehending 'that it exists *partes extra partes*' (Merleau-Ponty 2012: 75). And the ability to think about oneself in this way requires some awareness of oneself as being part of the objective world.

While it's possible for subjects with self-consciousness to achieve this, it's unclear whether minimal subjects, which are capable of thinking in the *de se* way but incapable of making these thoughts explicit, can think of themselves as an object. They're given to themselves only as the experiencing and interacting origin of the world and as such aren't a proper part of their world. While they might experience their own bodies visually, such an experience shouldn't be described as an experience *as of* an external object in the world. Describing it in such a way would require that the subject also apprehends some relations between her own physical body and other objects in the world. But this requires the ability to think about oneself in a self-conscious way. As such, we should say that only self-conscious subjects are subjects and objects at the same time, depending on how they think about themselves. Other subjects are merely given to themselves first-personally.

A.4.2 *Perspectives and submarines*

Some issues about the notion of perspectival space and its relation to egocentric space could use more discussion. The first concerns the seemingly clear notion of a perspective altogether. As a matter of fact, it's far from obvious whether we can clearly define perspectival space as a distinct kind. Because the distinction between allocentric, egocentric and perspectival space is drawn on an epistemological level, we might realise that the theoretical distinctions don't correspond to real

epistemic differences. For instance, Bennett (2009) intends to distinguish different varieties of visual perspectives. One main problem with this attempt are the different epistemic imports that are relevant to the varieties. As such, certain kinds of perspectives already make reference to possible ways for the subject to interact with the world—bringing it close to egocentric thinking.

This is closely connected to the question whether there are objective perspectival facts—a question that is relevant both in the philosophy of perception and in the philosophy of time (cf. Le Poidevin 2007: ch. 3). The suggestion is that we draw the distinctions above on a purely epistemological level in terms of how one and the same objective space is represented. A. W. Moore (1997) argues strongly for this way of understanding the distinction between absolute and perspectival thinking. However, at the same time, we need to do justice to the idea that some perspectival relations are perfectly objective (A.4.5). I suggest we retain the best of both worlds. As such, there's just one objective space which is thought about either in perspectival, egocentric, or allocentric terms. At the same time, this objective space can be said to incorporate 'perspectival' facts. This is because there are relations which obtain between the objects in the world. And some of these relations can be thought of in different ways depending on where the subject is located relative to the related objects.

A second issue concerns the necessity of egocentric thinking to account for the nature of *de se* thinking. Why shouldn't perspectival thinking be enough? We can find several different arguments from different theoretical projects supporting this claim. Next to the arguments presented in this book, we find Perry arguing that our behaviour is intrinsically connected to this kind of agent-relative knowledge: 'However complex our lives are, everything we do comes down to performing operations on the objects around us—objects in front of us, behind us, above us; objects we are holding; objects we can see. … Practical knowledge then, the knowledge that enables us to do things, forms a structure at whose base is information about the objects that play relatively basic agent-relative roles in our lives' (Perry 1998: 85). Further support comes from theories of situated cognition and enactivism (Rowlands 1999, 2010; Thompson 2007).

Finally, we can use a well-known case discussed by Evans to give additional support to the distinction between egocentric and perspectival

thinking. Evans wonders about the meaning of 'here'-thoughts in the context of thinking about a place from a certain perspective. In the example, there's a remote controlled submarine floating around the sea bed. The operator is located at the surface in a ship from where she can see the world from the submarine's point of view. She comes to think—on the basis of the video—things like *It's mucky here* or *What do we have here?* Now, is she thinking about the sea bed in egocentric terms? In other words, does her use of the concept 'here' refer to the centre of egocentric space? Evans argues that we don't want to say this in the normal case, but that there are possible situations in which it makes sense to say that the operator's 'here'-thoughts refer to the centre of egocentric space—i.e. the location of the submarine.

He argues that if the subject in this peculiar situation 'knows that the information does not concern her immediate environment, she will not locate the place in egocentric space, and so some other mode of identification will be in question. She will think of the place as *where the submarine is*, or *where these pictures are coming from*' (Evans 1982: 165). In such a case, we would say that our subject thinks about the sea bed in a merely perspectival way without assuming to be located at the bottom of the ocean. However, Evans then goes on to imagine the subject being more immersed in her activity in the submarine. The subject doesn't move on the ship anymore and the smells and sounds in her immediate environment aren't of any direct importance anymore. In such a case, 'the centre of her world would be down on the sea bed, and her utterances of "here" and "this" could go direct to their objects without the need for conceptual supplementation' (Evans 1982: 167). This move from merely perspectival to egocentric thinking is tantamount to a move to thinking of the objects in terms of one's possibilities of interaction.

A.4.3 *Campbell on egocentric space*

There's some connection between my proposal of egocentric thinking as being constituted through a subject's assumed possibilities of interaction and Campbell's notion of egocentric space. A chief difference is the claim that we can't give an accurate definition of why thinking in egocentric terms provides immediate reasons for the subject to act in a certain way. Accordingly, Campbell thinks that 'egocentric axes

are "immediately" used to direct action. It may be that no very precise definition can be given of this notion of immediate use, and that the notion of an egocentric reference frame must to this extent remain a rough and intuitive one' (Campbell 1994: 15–16).

In contrast to Campbell, I think there's a way to explain the notion of immediate use. We've already made steps into that direction in the context of characterising the functional approach to *de se* thinking in terms of 'direct importance' (A.2.9). There's no direct translation of these ideas and arguments to an explanation of why egocentric space can be 'immediately' used to direct action. At the same time, the connections are strong enough to warrant the claim that there's a possibility of going beyond a 'rough and intuitive' understanding.

A.4.4 *What kind of assumptions?*

The idea that egocentric thinking and, by extension, *de se* thinking have to be characterised and individuated in terms of the subject's assumed possibilities of interaction with the objects thereby thought about needs some clarification. Before attending to three more specific questions, it's important to illustrate the connection to other related phenomena. The first one is Campbell's (1994) argument that reasoning in the first person neither depends on the mere metaphysical identity of the reference of the various uses of the first person concept nor on an explicit assumption in the form of a first person judgement. Rather, the subject trades on the identity in question in her reasoning. The assumptions about interaction, which are in play in egocentric thinking, are similar to this kind of trading on identity.

The second connection is to Lucy O'Brien's *agency account* of *de se* thinking. There, she argues that the capacity for self-conscious self-reference depends on the awareness of being an agent. However, this kind of awareness needn't be conceptualised: 'Agent's awareness is a form of awareness that a subject has of her own actions, and that precedes her capacity to conceptualize it. It is a form of non-conceptual awareness' (O'Brien 2007: 88). Subjects who think in the *de se* way have to be aware of the fact that their active mental states and their behaviour are under their rational control. But this doesn't require the subject to actively engage with her judgements and beliefs. Rather, a subject has agent's awareness insofar 'as the thoughts of a subject are

objects of her rational agency—subject to her rational responsibility, revision and acceptance' (O'Brien 2007: 92). Again, the assumptions about interaction are similar in that they manifest themselves in a subject's disposition to believe and act in a certain way.

With these preliminary remarks in the background, it'll be easier to give answers to the following three questions: (i) Why does egocentric space need to be characterised in terms of activity? (ii) What's the nature of these assumptions? Are they judgements, expectations, representational? (iii) What does it mean for a thought to be constituted by an assumption about practical abilities?

The philosophical tradition is full of arguments that purport to draw a conceptual connection between visual space and perception on the one hand and bodily activity on the other (cf. Smith 2014). However, there are different ways of spelling out this constitutive relationship, ranging from merely empirical and contingent to definitional or conceptual. The important point for the view advocated in this book can be put in the following way. For a subject to be capable of knowing how to interact with the objects in her world, she needs to think of these objects in a way which puts them into a directly engaging relation to her own acting body. As such, the way the subject interacts with the world through her body is constitutive of the egocentric space that's particular of that subject. Hence, activity is constitutive of the nature of egocentric space not because that space is a space of activity itself. Rather, thinking about the world egocentrically is intrinsically tied to the lived body, which is usually understood in terms of certain possible ways of interaction.

If we take the connections to Campbell's and O'Brien's arguments sketched above into account, the answer to the second question should become more apparent. First off, the subject's assumed possibilities of interaction can't be judgements since that would lead to an infinite regress as we saw in the case of reasoning in the first person. Indeed, these assumptions shouldn't be properly characterised as occurrent mental states at all. They're manifest in what a subject takes herself to be capable of doing on the basis of thinking about the world in this or that way. In this sense, they're similar to a form of dispositional expectation such as when the subject expects her room to be the same upon turning off the light. There's no explicit judgement or consciously held expectation involved in the subject's assumption concerning the

relative location of her bed. But the fact that she securely moves about her room in darkness shows that she thinks that the bed is still in the same place.

This is also relevant to the question whether we should think of these assumptions as representations. There are two arguments against this way of understanding the relevant notion of assumption that's in play. First off, they're supposed to ground the capability of *de se* thinking. But, if they're representational, then what's being represented—the possibilities of interaction—needs to be apprehended in a *de se* way in order to be taken as directly important to the subject which leads to a new infinite regress. Thus, they can't be representations. Secondly, thinking of them as representational gets the phenomenology and epistemology wrong. If I'm under the illusion that I might catch the ball, I assume some possibility of interacting with the ball. However, I might, as in the case of virtual reality, look through the illusion and consciously judge that I can't catch the ball. If the assumption were a representation, I should now form a new representation which deletes or overrides the former one. However, this doesn't seem right. Rather, my new conscious judgement merely trumps my illusory assumption. Hence, the relevant assumptions have different phenomenal and epistemic properties than representational states.

The answers to the first two questions now provide us with enough material for a reply to the third one. A thought being constituted by an assumption about practical abilities means that there are certain dispositions to make judgements according to these assumptions and behave as if there's a certain way of interacting with the world. Thinking of the glass as *being in reach*—understood as involving the assumption that I can grab it by moving in a particular way—implies that I would judge the glass as not being in my hand, that I would move my hand in a particular way, and so on. Hence, the thought is embedded in a certain network of possible and actual actions, possible and actual judgements and beliefs. And this network is shaped by the possibilities of interaction with the objects which the subject assumes.

A.4.5 *Egocentric space and objectivity*

There are two important questions about egocentric space which might require some further discussion. First of all, we want to know in what

sense egocentric space constitutes a kind of 'space' which needs to be distinguished from allocentric space. And secondly, we might wonder about the objectivity of egocentric space. Concerning the first question, we already saw Soldati (1998) and A. W. Moore (1997) arguing that egocentric space shouldn't be contrasted with objective space in metaphysical terms but rather in representational terms. But, at the same time there's a sense to the idea that perspectival facts can be perfectly objective. Let me now give some more support to the claim that egocentric space is characterised in representational terms while still being in some sense objective.

Again, Evans provides us with a nice argument for the claim that there's just one space which can be thought about either egocentrically or allocentrically:

> Notice that when I speak of information 'specifying a position in egocentric space', I am talking not of information about a special kind of space, but of a special kind of information about space—information whose content is specifiable in an egocentric spatial vocabulary. It is perfectly consistent with the *sense* I have assigned to this vocabulary that its terms should *refer* to points in a public three-dimensional space. (Indeed I shall be claiming that that is what they refer to, if they refer to anything at all.)
>
> Evans 1982: 157

Evans's point has to do with reference and the possibility to communicate and share thoughts. If we want to say that Alpha's belief *I'm being attacked by a bear* and Beta's belief *Alpha is being attacked by a bear* are about one and the same object, we should treat their beliefs as referring to one and the same public space. But if that's so, then Alpha's egocentric way of thinking about the bear can't be understood as thinking about a different kind of space. For, if that were the case, we would have to explain how her belief could possibly refer to the same object as Beta's belief. Hence, egocentric space has to be characterised in terms of a 'mode of presentation' of objective space.

The question whether egocentric thinking can be objective is closely connected. Egocentric space and perspectival space are distinguished by the question whether the subject takes herself to be capable of acting from the centre of the respective space. Hence, we can try to tackle the

question of objectivity via thinking about the possibility of perspectival facts. The relations which are constitutive of perspectival space can be understood in a perfectly objective or intersubjective way. What perspectival thinking claims is conditional on a certain point of view—i.e. a certain centre in objective allocentric space—but not conditional on a specific subject occupying that point of view. In this sense, K2 is to the right of Gasherbrum II when viewed from Gasherbrum I for every subject who takes up that particular perspective. There's no variation here.

This insight can be traced back at least to Thomas Reid's *An Inquiry into the Human Mind on the Principles of Common Sense* (1764), where he argues for the objectivity of what he calls the visible figure of an object—the object as presented to the senses with all its geometrical properties. He argues that 'the visible figure of all bodies will be the same with that of their projection upon the figure of a hollow sphere, when the eye is placed in the centre' (Reid 1764: 6.7, 218–219). The projection of the surface on the inside of a hollow sphere seems to be a perfectly objective figure. If that's the case, then the visible figure—how an object appears perspectivally—is perfectly objective. Accordingly, Reid claims that the visible figure belongs to the category of real figures:

> To what category of beings does visible figure then belong? … The different positions of the several parts of the body with regard to the eye, when put together, make a real figure, which is truly extended in length and breadth, and which represents a figure that is extended in length, breadth, and thickness.
>
> Reid 1764: 6.8, 225–226

A.4.6 *Is egocentric thinking de se?*

Let's imagine you accept some conceptual connection between thinking about space egocentrically and *de se* thinking. Then you might still wonder about two theses that resonate in the description of egocentric space. First off, do we want to say that a subject thinks in the *de se* way whenever she interacts with objects in the world? Secondly, is every case of *de se* thinking a case of self-consciousness? For the purposes

of the book, I accept a version of the first thesis and firmly reject the second. Here's the reasoning behind it.

As we'll see in the context of discussing some examples of *de se* thinking, even cases of egocentric thinking about other *objects* involves the self-ascription of some properties. Thinking that there's food to the left doesn't strike us as a very *de se* way of thinking. After all, we're thinking about the food and not about ourselves. Hence, the intentional object of the thought isn't the subject. With Perry (1986) and Récanati (2007), we might say that these thoughts 'concern' the subject but they aren't 'about' the subject. Nonetheless, the subject can only think about the food as being in a certain location in egocentric space if she thinks of it in relation to *herself*. And this way of thinking of the object involves the lived body and grounds a certain self-ascription of a property. Hence, all thinking about objects in egocentric terms involves self-ascription of a property. Therefore, it involves *de se* thinking. Thinking about an object egocentrically is only possible when put into connection with our own assumed possibilities of interaction.

On the other hand, self-consciousness should be seen as the kind of ability that subjects reveal in the classic mirror test (cf. Gallup 1970, 1977). As such, it can be understood as a highly sophisticated cognitive ability which isn't exhibited by all subjects capable of *de se* thinking. What many philosophers call primitive self-consciousness or non-conceptual self-consciousness has to be explained via the ability to think in the *de se* way without assuming that it already amounts to proper self-consciousness.

A.4.7 *The lived body, the body image, and the body schema*

The distinction between the lived body and the physical body is related, but not identical to the distinction between the body image and the body schema. Both the body image and the body schema are built through the subject's intentional states on her body. They take the subject's body as an intentional object. However, the body image involves a set of perceptions, representations, beliefs and attitudes *towards* the own body while the body schema is more implicit in kind. Thus, the body schema supposedly works on a more primitive and functional level than the partially explicit attitudes towards one's own body which are constitutive of the body image. Gallagher and Cole (1995), who

established this conceptual distinction, take the body image to comprise a subject's perceptual experience and conceptual understanding of her own body paired with different emotional attitudes towards the body. In contrast, the body schema doesn't have the 'status of a conscious representation or belief' (Gallagher and Cole 1995: 371). It is integrated with the subject's environment and is determined through 'prepersonal, anonymous processes' (Gallagher and Cole 1995: 373) which serve to establish the assumed boundaries of one's own behavioural body. As such, it may incorporate objects distinct from the actual physical body—such as hammers, walking sticks, artificial limbs, and so on. The body schema is very much tied to direct movement and interaction and can even consist of several body schemas which are subsumed under one general heading (Gallagher 2005: 24, fn 5).

Neither the body schema nor the body image are anything like the objective physical body. Both involve some form of intentional stance towards one's own body. What about the relation to the lived body? Being constituted through partially explicit intentional attitudes to one's own body, the body image has to be sharply distinguished from the lived body, which enables intentional attitudes towards one's own body and thus can't itself be constituted thusly. At the same time, one could be lead to think that the body schema is very closely related to the lived body. However, the two are not identical at all. While the body schema involves the system of motor abilities which is very much crucial in the case of the lived body, the concept of the lived body is located on an almost exclusively phenomenological and epistemic level. The lived body is neither consciously nor unconsciously *represented* at all. For otherwise, the subject would have to think of the lived body *as* her own. But this, of course, would require some prior knowledge of who she herself is—a prerequisite that the lived body is exactly designed to fulfil.

There's certainly a connection between egocentric space and the body schema because the latter is determined with regard to the motor capacities of the subject and egocentric space has to be characterised in terms of activity. However, we can't take the body schema to be the origin of egocentric space. This is because the kind of bodily awareness that's typical of the lived body is more fundamental than the implicitly represented body schema. As Soldati (1998: 144) argues: 'More is required for egocentric space than the centring of a spatial frame on the subject's body. What is needed, it might be argued, is some *non-obser-*

vational knowledge of one's own body, as it is generally said to be given by bodily awareness'. And this non-observational knowledge can't be provided by the body schema because the subject would have to think of a particular body schema as *hers*. But this doesn't amount to a way of thinking that's potentially identification-free.

A.4.8 *Is the lived body first-personal?*

An account of *de se* thinking which is conceptually quite close to the defended idea involving the lived body is Lucy O'Brien's *agency account* which she defends in her *Self-Knowing Agents* (2007). On her account, a subject is capable of self-conscious reference to herself using the first-person pronoun or concept via her understanding the self-reference rule—a user of the word 'I' or the concept *I* refers necessarily to herself—only if she thereby exhibits what she calls 'agent's awareness' of her use. In her use, she has to be aware of being the agent in order to become aware of the fact that she's thinking about herself. Without such a primitive awareness of being the rational agent of her thought or utterance, the subject wouldn't be able to know that she thinks about *herself* through using the first-person concept (A.4.4).

O'Brien's arguments for the necessity of primitive agent's awareness mirror the arguments presented for the necessity of primitive self-ascription. Furthermore, the notion of agent's awareness is somewhat close to the concept of the lived body. Both involve a primitive kind of knowledge about one's own behaviour and possibilities of interaction. However, the two are not identical. Agent's awareness has much to do with a subject's rational control over her mental and behavioural life. For instance, a particular experience of anger, despite being passive and overcoming oneself, is experienced as one's *own* anger in virtue of the fact that the subject is capable of fitting that anger into a picture of a rational agent. For instance, the anger has a reasonable cause, is appropriate and doesn't just come out of thin air. By contrast, a case of the experience of thought insertion—i.e. the experience of a mental state which isn't accepted as one's own—doesn't exhibit agent's awareness and thus isn't taken as one's own thought. There is no rational control over the thought which just appears to be 'in one's head'.

In contrast, the lived body operates on a much more basic level without the necessity for expansive rationality of the subject. A sub-

ject with a lived body needn't exert rational control over her mental state in order to be capable of thinking about herself. This also contrasts O'Brien taking agent's awareness to be essentially *not* first-personal. For an ability that is supposed to ground our ability to think in the *de se* way, this is quite surprising. However, she assumes that agent's awareness being first-personal would require a further explanation of that feature. She reasons that our use of the word 'I' is first-personal because of the nature of the self-reference rule and agent's awareness. So, we can't take the explanans of the first-personality of self-conscious self-reference itself to be first-personal.

However, this worry is mistaken. In fact, what requires explanation isn't the first-personality of our ability to use the word 'I' in order to express our *de se* mental states. What requires explanation is that something which is governed by purely third-personal rules is capable of expressing irreducibly first-personal attitudes. And the idea that the lived body is the essential first-personal basis for *de se* thinking is exactly an explanation of that. It's the rock bottom first-personality of the lived body which makes it the case that subjects are able to have a first-personal stance on the world and themselves *at all*.

A.4.9 *More on the lived body*

The contrast between the physical and the lived body introduces some problems concerning the spatial nature of the latter. We might ask whether the lived body—as the origin of egocentric space—has an extension and in what sense fingers, hands, or cheeks are 'parts of' the lived body. Admittedly, we can't describe the geometrical origin of objective space in divisible terms. But the lived body isn't a geometrical origin, it's a phenomenological one. Thus, here's a proposal: The physical and the lived body aren't part of distinct metaphysical realms. In this sense, then, the two can overlap spatially or even have the same spatial extension. For usual subjects, their physical hands occupy the same part of objective space as their lived hands. However, cases such as the rubber hand illusion, phantom or alien limbs show us that the lived body is much more malleable and susceptible to transformation.

We can metaphorically imagine the lived body as a kind of body-shaped overlay that's determined by the two-way relation between experience and behaviour. If a subject assumed that there's a tall horn

on top of her head with which she can receive information from outer space and that can send information to extraterrestrials, then that physically non-existing horn should be understood as a part of her lived body. In our metaphor, the subject's 'overlay' is comprised of hands and feet, a face, a tongue, and also the tall horn on top of the head. In this sense then, there's a way to comprehend the lived body as located in physical space. However, the extension is determined phenomenologically and not physically.

Another question concerns the claim that the lived body is free from identification. Is there a way to make this claim more precise? I suggest two clarifications: First, any thought that's grounded in the ascription of a property to the lived body is identification-free. Why? Because 'we do not need to identify our body as one among those experienced' (Soldati 1998: 144). Or, as Martin (1993: 209) argues: 'Bodily awareness is such because its proper and sole object is one's own body and not any other occupant of the objective world'. For thoughts that involve primitive self-ascription, there's no question about which object one ascribes a property to in sense similar to introspection always being about one's own mind and not any other mind. Along the same lines, Merleau-Ponty (2012: 93) argues that 'I observe external objects with my body, I handle them, inspect them, and walk around them. But when it comes to my body, I never observe it itself'.

This leads to the second clarification against a possible misinterpretation of the claims above. The fact that there's merely one lived body to which subjects ascribe properties doesn't imply that there's no possibility to think about one's own body as one amongst others. It's perfectly possible to experience one's body 'as an object in a world which can contain other objects' (Martin 1993: 209). However, this way of thinking of the physical body requires an awareness of oneself as being part of the objective world. Hence, it requires the ability to leave one's first-person perspective and apprehend oneself as located in allocentric space (A.4.1).

A.4.10 *Is de se thought without a body possible?*

There's an obvious objection around the corner here. Aren't there examples of possible disembodied subjects—such as the souls of religious people in the afterlife, Descartes's pure thinker, or something like a

mind that's only possible of calculation—which are capable of thinking about themselves in the *de se* way? It seems rather straightforward that such subjects present direct counterexamples to the lived body account. Therefore, let me quickly discuss two strategies with which we could reply to such an objection.

The first concerns the intelligibility of the objection either on empirical or conceptual grounds. We could question the existence of these disembodied subjects. After all, believing something to be true doesn't make it true. In order to avoid the inevitable squabble over such a reply, we can also question whether it makes sense to attribute these subjects with the ability to think in the *de se* way. As argued in the main text, the *Cogito* argument doesn't require the acceptance of disembodied subjects. Rather, it tells us that the knowledge of our own existence is epistemically independent from any kind of observational knowledge about our bodies. But this is perfectly consistent with the lived body account. However, this still leaves us with the supposed possibility of a soul in Heaven thinking *God loves me*. How should we respond to this?

I propose a second strategy that doesn't commit us to any specific metaphysical claims about the existence of souls or the capability of these subjects to think in the *de se* way. Rather, it tries to show a way in which these subjects could be understood as being constituted by a lived body in the purely phenomenological sense. As long as this subject can be said to experience the world in terms of assumed possibilities of interaction with the 'objects' in her world, it could be said to be a lived body. Since the lived body isn't identical to a physical body, it being disembodied in the physical sense is a conceptual possibility.

How plausible is that? I don't know. But, I'll present one very hypothetical far-fetched way of making sense of a lived body—including some sense of egocentric space—without any interaction with physical objects. Imagine a subject without a body which is capable of calculating. However, that subject experiences numbers as abstract objects which are located in her egocentric space. How so? Let's think of a conception of arithmetic as we find it in Kant's *Critique of Pure Reason* and *Prolegomena to Any Future Metaphysics*, where he argues that 'even arithmetic forms its concepts of numbers through successive addition of units in time, but above all pure mechanics can form its concepts of motion only by means of the representation of time' (Kant 2004:

4.283). Now, against this background, we might imagine our thinking subject as thinking about numbers in terms of how long it takes her to represent them. Thus, in order to 'interact' with the number 5, she thinks of it as taking two units longer to reach than the number 3. In this way, her world of numbers can be said to be represented along the stretch of time which it takes her to reach the relevant objects in representation. This could be described as a kind of egocentric way of thinking about abstract numbers. And our subject thinks of this world in terms of how she could interact with the various objects it contains. This would allow us to say that there's a lived body of some sorts involved. Admittedly, this is somewhat speculative, but it could be one way of making sense of purely thinking subjects.

A.4.11 *Implicit de se beliefs*

Some accounts of first person thought make a more systematic distinction between *implicit* and *explicit de se* beliefs (Musholt 2015; Récanati 2007, 2009). The latter contain some element or constituent in the content that makes them explicitly about the thinking subject. For example, the occurrence of the first person concept would be a constituent that makes the belief explicitly *de se*. What about implicit *de se* beliefs? Récanati explains:

> Thoughts that are implicitly *de se* involve no reference to the self at the level of content: what makes them *de se* is simply the fact that the content of the thought is evaluated with respect to the thinking subject. The subject serves as 'circumstance of evaluation' for the judgment, rather than being a constituent of content.
>
> Récanati 2009: 258

It's not quite clear whether cases of agent-relative knowledge (cf. Perry 1998) should be properly analysed in terms of implicit *de se* thinking. They fulfil the essential criteria above in that their content needs to be evaluated for its satisfaction with respect to the thinking subject. However, because their intentional object isn't the thinking subject, it's questionable whether we want to give them the status of *de se* states.

A.4.12 *Affordances and de se thinking*

A common way to put this point is to say that many of our beliefs about the world around us are guided by what's called the perception of *affordances*. Objects in the world have certain characteristics that are relative to subjects engaging with them. For instance, grass is eatable for many ruminant animals like cows, giraffes, and kangaroos while for us it doesn't afford to be eaten. James Gibson, who is associated the most with the theory of affordances—despite having its roots in the much earlier Gestalt psychology—explains that affordances are properties of objects that are relative to the observing subject. They are relative in so far as the subject is of a certain kind herself. Grass affords different things to human beings than to giraffes. And an open window affords something different to a thief than to a freezing subject.

In this sense then, affordances aren't subjective, they're just relational properties of objects that depend on which subject engages with it. Accordingly, Gibson writes in his 'The Theory of Affordances' (1977):

> The concept of affordance is derived from these concepts of valence, invitation, and demand but with a crucial difference. The affordance of something does *not change* as the need of the observer changes. The observer may or may not perceive or attend to the affordance, according to his needs, but the affordance, being invariant, is always there to be perceived. An affordance is not bestowed upon an object by a need of an observer and his act of perceiving it. The object offers what it does because it is what it is. To be sure, we define *what it is* in terms of ecological physics instead of physical physics, and it therefore possesses meaning and value to begin with. But this is meaning and value of a new sort.
>
> Gibson 1977: 138–139

What it means for a subject to think of food as being 'in reach' is just for her to latch on to the affordance that the food item has in virtue of standing in a particular relation to the thinking subject. The relation of being in reach always obtains between that food item and subjects capable of reaching and within a certain distance. But, while

the affordance itself isn't subjective, it carries an immediate call for action to the subject with it. This is because a subject that perceives the affordance of an object has to think of that object as in reach *for herself*. Thus, perceiving an affordance carries direct relevance for action because affordances are such that they're determined in terms of what a subject can do with a certain object.

A similar way of describing our thoughts about objects in egocentric space comes from enactivism about perception (Hutto and Myin 2012). For instance, Alva Noë explains this theory of perception, strongly inspired by Merleau-Ponty's concept of the lived body, thusly:

> Perceptual experience acquires content thanks to our possession of bodily skills. *What we perceive* is determined by *what we do* (or what we know how to do); it is determined by what we are *ready* to do. In ways I try to make precise, we *enact* our perceptual experience; we act it out.
>
> Noë 2004: 1

A.4.13 *Strawson on feature placing*

This way of putting things is derived from Peter Strawson's concept of *feature placing* that he first develops in his *Individuals* (1959). The idea is that some sentences and mental states don't designate an object and a property or predicate. Rather, they place a certain feature—such as warmth, rain, or colour—in one's environment. Accordingly, he distinguishes 'feature-universals' from properties: '*Snow*, *water*, *coal* and *gold*, for example, are general kinds of stuff, not properties or characteristics of particulars; though *being made of snow* or *being made of gold* are characteristics of particulars' (Strawson 1959: 202).

How is this relevant for us? We could understand primitive self-ascription as placing a certain feature 'in' the lived body. Correspondingly, such a self-ascription wouldn't involve an object. A subject that has the *de se* thought *My back is horizontal* merely places a certain feature in her lived body without picking herself out as the object of that thought. It's not quite clear if it's possible to translate Strawson's language of feature placing directly to the lived body account. The reason is that I didn't develop a mature theory of properties to go along with the account—having argued that this isn't necessary to get the gen-

eral point across. Accordingly, the idea of feature placing might only be compatible with certain theories of properties and not with others. But we would need a clearer theory of properties to test this.

A.4.14 *Self-consciousness in nonhumans*

Discussions on the viability of the mirror test to detect self-consciousness in nonhumans tell us much about the complexity of empirical science. First off, it's quite unclear whether we can use a single criterion like the mirror test to determine whether some subject is self-conscious or not (Wüstholz 2015). Because self-recognition isn't a monolithic thing but might be 'conceived of as a gradual phenomenon' (Brandl 2016: 2), we need to look for an encompassing paradigm to understand self-consciousness. This is even more so, because the behaviour exhibited in mirror self-recognition can easily be explained in terms that don't make reference to self-consciousness (Wüstholz 2013).

REFERENCES

Anscombe, Gertrude Elizabeth Margaret

1975 'The First Person'. In *Mind And Language: Wolfson College Lectures 1974*. Oxford: Clarendon Press, pp. 45–64.

Bayne, Tim and Michelle Montague

2011 (eds.), *Cognitive Phenomenology*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199579938.001.0001.

Bennett, David J.

2009 'Varieties of visual perspectives'. *Philosophical Psychology*. 22, 3, pp. 329–352. DOI: 10.1080/09515080902970665.

Bermúdez, José Luis

1998 *The Paradox of Self-Consciousness*. MIT Press.

Bilgrami, Akeel

2012 'The Unique Status of Self-Knowledge'. In *The Self and Self-Knowledge*. Ed. by Annalisa Coliva. Oxford University Press, pp. 263–278. DOI: 10.1093/acprof:oso/9780199590650.003.0013.

Boghossian, Paul A.

2008 *Content & Justification*. Oxford University Press.

Botvinick, Matthew and Jonathan Cohen

1998 'Rubber hands "feel" touch that eyes see'. *Nature*. 391, p. 756. DOI: 10.1038/35784.

Brandl, Johannes

2014 'Die Entwicklung der Autorität der Ersten Person'. *Deutsche Zeitschrift für Philosophie*. 62, 5, pp. 937–962. DOI: 10.1515/dzph-2014-0061.

Brandl, Johannes

2016 'The puzzle of mirror self-recognition'. *Phenomenology and the Cognitive Sciences*, pp. 1–26. DOI: 10.1007/s11097-016-9486-7.

Brandom, Robert B.

1998 *Making it Explicit*. Harvard University Press.

Brentano, Franz

2009 *Psychology from an Empirical Standpoint*. Routledge.

Bringhurst, Robert

2004 *The Elements of Typographic Style*. 3rd ed. Hartley & Marks.

Burge, Tyler

1998 'Reason and the First Person'. In Wright et al. (1998), pp. 243–270. DOI: 10.1093/0199241406.003.0009.

2007 *Foundations of Mind*. Clarendon Press.

Campbell, John

1994 *Past, Space, and Self*. MIT Press.

2012 'Lichtenberg and the *Cogito*'. *Proceedings of the Aristotelian Society*. 112, 3, pp. 361–378. DOI: 10.1111/j.1467-9264.2012.00341.x.

Camus, Albert

1955 *The Myth of Sisyphus and Other Essays*. New York: Alfred A. Knopf.

Cappelen, Herman and Josh Dever

2013 *The Inessential Indexical*. Context and Content. Oxford University Press. DOI: 10.1093/acprof:oso/9780199686742.001.0001.

Carnap, Rudolf

1950 'Empiricism, Semantics, and Ontology'. *Revue Internationale de Philosophie*. 4, pp. 20–40.

Castañeda, Hector-Neri

1999a 'Indicators and Quasi-Indicators'. In Castañeda (1999d), pp. 61–88.

1999b 'Philosophical Method and the Direct Awareness of the Self'. In Castañeda (1999d), pp. 96–142.

1999c 'Self-Consciousness, Demonstrative Reference, and the Self-Ascription View of Believing'. In Castañeda (1999d), pp. 143–179.

1999d *The Phenomeno-Logic of the I*. Ed. by James G. Hart and Tomis Kapitan. Indiana University Press.

Chalmers, David J.

2011 'Propositions and Attitude Ascriptions: A Fregean Account'. *Noûs*. 45, 4, pp. 595–639. DOI: 10.1111/j.1468-0068.2010.00788.x.

Chisholm, Roderick M.

1976 *Person and Object: A Metaphysical Study*. Open Court.

1981 *The First Person: An Essay on Reference and Intentionality*. University of Minnesota Press.

Chudnoff, Elijah

2015 *Cognitive Phenomenology*. New Problems of Philosophy. Routledge.

Cocchiarella, Nino B.

2007 *Formal Ontology and Conceptual Realism*. Springer. DOI: 10.1007/978-1-4020-6204-9.

Coliva, Annalisa

2003 'Moore's Proof of an external world. Just begging the question'. In *Proceedings Wittgenstein Symposium Kirchberg XI*. Ed. by Winfried Löffler and Paul Weingartner. Kirchberg am Wechsel, pp. 91–96.

2016 *The Varieties of Self-Knowledge*. Palgrave Macmillan. DOI: 10.1057/978-1-137-32613-3.

Descartes, René

2008 *Meditations on First Philosophy*. Ed. and trans. by Michael Moriarty. Oxford World's Classics. Oxford University Press.

Doyle, Arthur Conan

1892 *Adventures of Sherlock Holmes*. Harper & Brothers, Franklin Square.

Dummett, Michael

1973 *Frege: Philosophy of Language*. Harper & Row.

Egan, Andy

2004 'Second-Order Predication and the Metaphysics of Properties'. *Australasian Journal of Philosophy*. 82, 1, pp. 48–66. DOI: 10.1080/713659803.

Evans, Gareth

1982 *The Varieties of Reference*. Oxford University Press.

1985 'Understanding Demonstratives'. In *Collected Papers*. Clarendon Press, pp. 291–321.

Feit, Neil

2008 *Belief About the Self: A Defense of the Property Theory of Content*. Oxford University Press. DOI: 10.1093/acprof:oso/9780195341362.001.0001.

Feyerabend, Paul

1993 *Against Method*. 3rd ed. Verso.

Frege, Gottlob

1953 *The Foundations of Arithmetic*. Trans. by John Langshaw Austin. Harper & Brothers.

1956 'The Thought: A Logical Inquiry'. *Mind*. 65, 259, pp. 289–311. DOI: 10.1093/mind/65.1.289.

Gallagher, Shaun

2005 *How the Body Shapes the Mind*. Oxford University Press. DOI: 10.1093/0199271941.001.0001.

Gallagher, Shaun and Jonathan Cole

1995 'Body Image and Body Schema in a Deafferented Subject'. *Journal of Mind and Behavior*. 16, 4, pp. 369–390.

Gallup, Gordon G.

1970 'Chimpanzees: self-recognition'. *Science*. 167, 3914, pp. 86–87. DOI: 10.1126/science.167.3914.86.

1977 'Self recognition in primates: A comparative approach to the bidirectional properties of consciousness.' *American Psychologist*. 32, 5, pp. 329–338. DOI: 10.1037/0003-066X.32.5.329.

Gibson, James J.

1977 'The Theory of Affordances'. In *Perceiving, Acting, Knowing*. Ed. by Shaw and Bransford. Lawrence Erlbaum Associates, pp. 127–143.

Hemingway, Ernest

2014 *The Sun Also Rises*. Scribner.

Holton, Richard

2015 'Primitive Self-Ascription: Lewis on the *De Se*'. In *A Companion to David Lewis*. Ed. by Barry Loewer and Jonathan Schaffer. Wiley-Blackwell, pp. 399–410.

Husserl, Edmund

1973 *Zur Phänomenologie der Intersubjektivität. Erster Teil.* Ed. by Iso Kern. Martinus Nijhoff.

1983 *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*. Trans. by Fred Kersten. Martinus Nijhoff. Vol. 1.

2001 *Logical Investigations*. Trans. by J. N. Findlay. Routledge. Vol. 1.

Hutto, Daniel D. and Erik Myin

2012 *Radicalizing Enactivism*. MIT Press. DOI: 10.7551/mitpress/9780262018548.001.0001.

Jackson, Frank

1999 'Non-Cognitivism, Normativity, Belief'. *Ratio*. 12, 4, pp. 420–435. DOI: 10.1111/1467-9329.00102.

Kalckert, Andreas and H. Henrik Ehrsson

2012 'Moving a rubber hand that feels like your own: a dissociation of ownership and agency'. *Frontiers in Human Neuroscience*. 6, pp. 1–14. DOI: 10.3389/fnhum.2012.00040.

Kant, Immanuel

1998 *Critique of Pure Reason*. Ed. and trans. by Paul Guyer and Allen W. Wood. The Cambridge Edition of the Works of Immanuel Kant. Cambridge University Press.

2004 *Prolegomena to Any Future Metaphysics*. Trans. by Gary Hatfield. Cambridge Texts in the History of Philosophy. Cambridge University Press.

Kapitan, Tomis

2016 'Indexical Duality: A Fregean Theory'. *Rivista Internazionale di Filosofia e Psicologica*. 7, 3, pp. 303–320. DOI: 10.4453/rifp.2016.0033.

Kaplan, David

1989 'Demonstratives'. In *Themes From Kaplan*. Ed. by Joseph Almog, John Perry and Howard Wettstein. Oxford University Press, pp. 481–563.

Le Poidevin, Robin

2007 *The Images of Time: An Essay on Temporal Representation*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199265893.001.0001.

Levine, Steven

2016 'Sellars and Nonconceptual Content'. *European Journal of Philosophy*. 24, 4, pp. 855–878. DOI: 10.1111/ejop.12127.

Lewis, David

1966 'An Argument for the Identity Theory'. *The Journal of Philosophy*. 63, 1, pp. 17–25. DOI: 10.2307/2024524.

1979 'Attitudes *De Dicto* and *De Se*'. *Philosophical Review*. 88, 4. Reprinted in Lewis (1983b): 133–156, pp. 513–543. DOI: 10.2307/2184843.

1983a 'New work for a theory of universals'. *Australasian Journal of Philosophy*. 61, 4, pp. 343–377. DOI: 10.1080/00048408312341131.

1983b *Philosophical Papers*. Oxford University Press. Vol. 1.

1986 *On the Plurality of Worlds*. Blackwell Publishers.

1998 'Index, context, and content'. In *Papers in Philosophical Logic*. Cambridge University Press, pp. 21–44.

Liao, Shen-yi

2012 'What are centred worlds?' *The Philosophical Quarterly*. 62, 247, pp. 294–316. DOI: 10.1111/j.1467-9213.2011.00042.x.

Lichtenberg, Georg Christoph

1990 *Aphorisms*. Trans. by R. J. Hollingdale. Penguin Books.

Magidor, Ofra

2015 'The myth of the de se'. *Philosophical Perspectives*. 29, 1, pp. 249–283. DOI: 10.1111/phpe.12065.

Margolis, Eric and Stephen Laurence

2007 'The Ontology of Concepts—Abstract Objects or Mental Representations?' *Noûs*. 41, 4, pp. 561–593. DOI: 10.1111/j.1468-0068.2007.00663.x.

Martin, Michael

1993 'Sense modalities and spatial properties'. In *Spatial Representation*. Ed. by Naomi Eilan, Rosaleen McCarthy and Bill Brewer. Clarendon Press, pp. 206–218.

Marx, Karl and Frederick Engels

1975 *Collected Works*. International Publishers. Vol. 5.

McDowell, John

1984 '*De Re* Senses'. *The Philosophical Quarterly*. 34, 136, pp. 283–294. DOI: 10.2307/2218761.

1996 *Mind and World*. Harvard University Press.

Melville, Herman

2003 *Moby-Dick, or The Whale*. Penguin Classics.

Merleau-Ponty, Maurice

2012 *Phenomenology of Perception*. Trans. by Donald A. Landes. Routledge.

Milne, Alan Alexander

1926 *Winnie-the-Pooh*. 1st ed. Methuen & Co. Ltd.

Moore, Adrian William

1997 *Points of View*. Oxford University Press.

Moore, George Edward

1939 'Proof of an External World'. *Proceedings of the British Academy*. 25, pp. 273–300.

Musholt, Kristina

2013 'Self-Consciousness and Nonconceptual Content'. *Philosophical Studies*. 163, 3, pp. 649–672. DOI: 10.1007/s11098-011-9837-8.

2015 *Thinking about Oneself*. MIT Press. DOI: 10.7551/mitpress/9780262029209.001.0001.

Noë, Alva

2004 *Action in Perception*. Representation and Mind. MIT Press.

O'Brien, Lucy

2007 *Self-Knowing Agents*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199261482.001.0001.

Ovid

1922 *Metamorphoses*. Trans. by Brookes More. Boston, Cornhill Publishing Co.

Peacocke, Christopher

2012a 'Descartes Defended'. *Proceedings of the Aristotelian Society*. 112, 3, pp. 109–125. DOI: 10.1111/j.1467-8349.2012.00210.x.

2012b 'Explaining *De Se* Phenomena'. In *Immunity to Error through Misidentification: New Essays*. Ed. by Simon Prosser and François Récanati. Cambridge University Press, pp. 144–157. DOI: 10.1017/CBO9781139043274.009.

2014 *The Mirror of the World*. Context and Content. Oxford University Press. DOI: 10.1093/acprof:oso/9780199699568.001.0001.

Perry, John

1977 'Frege on Demonstratives'. *Philosophical Review*. 86, 4, pp. 474–497. DOI: 10.2307/2184564.

1979 'The Problem of the Essential Indexical'. *Noûs*. 13, pp. 3–21. DOI: 10.2307/2214792.

1980 'A Problem About Continued Belief'. *Pacific Philosophical Quarterly*. 61, pp. 317–332.

1986 'Thought Without Representation'. *Proceedings of the Aristotelian Society*. 60, 1, pp. 137–152. DOI: 10.1093/aristoteliansupp/60.1.137.

1990 'Self-Notions'. *Logos*. 11, pp. 17–31.

1998 'Myself and I'. In *Philosophie in synthetischer Absicht*. Ed. by Marcello Stamm. Klett-Cotta, pp. 83–103.

2002 'The Self, Self-Knowledge, and Self-Notions'. In *Identity, Personal Identity, and the Self*. Hackett Publishing Company, pp. 189–213.

Pettit, Philip

2003 'Looks as Powers'. *Philosophical Issues*. 13, 1, pp. 221–252. DOI: 10.1111/1533-6077.00013.

Plato

1997 'Apology'. In *Complete Works*. Ed. by John M. Cooper. Hackett Publishing Company, pp. 17–36.

Prosser, Simon

2015 'Why Are Indexicals Essential?' *Proceedings of the Aristotelian Society*. 115, pp. 211–233. DOI: 10.1111/j.1467-9264.2015.00392.x.

Pryor, James

2004 'What's wrong with Moore's argument?' *Philosophical Issues*. 14, pp. 349–378. DOI: 10.1111/j.1533-6077.2004.00034.x.

Quine, Willard Van Orman

1953 'Two dogmas of empiricism'. In *From a Logical Point of View*. Harper & Row, pp. 20–46.

Ramsey, Frank Plumpton

1931 *The Foundations of Mathematics and Other Logical Essays*. Routledge.

Récanati, François

2007 *Perspectival Thought: A Plea for (Moderate) Relativism*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199230532.001.0001.

2009 '*De Re* and *De Se*'. *Dialectica*. 63, 3, pp. 249–269. DOI: 10.1111/j.1746-8361.2009.01194.x.

2012 *Mental Files*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199659982.001.0001.

Reid, Thomas

1764 *An Inquiry into the Human Mind on the Principles of Common Sense*. Edinburgh: printed for A. Millar, London, and A. Kincaid & J. Bell.

Rowlands, Mark

1999 *The Body in Mind*. Cambridge University Press. DOI: 10.1017/CBO9780511583261.

2010 *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*. MIT Press. DOI: 10.7551/mitpress/9780262014557.001.0001.

Sainsbury, Mark and Michael Tye

2012 *Seven Puzzles of Thought and How to Solve Them*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199695317.001.0001.

Shoemaker, Sydney

1968 'Self-Reference and Self-Awareness'. *The Journal of Philosophy*. 65, 19, pp. 555–567. DOI: 10.2307/2024121.

Siegel, Susanna

2011 *The Contents of Visual Experience*. Oxford University Press. DOI: 10.1093/acprof:oso/9780195305296.001.0001.

Smith, Joel

2014 'Egocentric Space'. *International Journal of Philosophical Studies*. 22, 3, pp. 409–433. DOI: 10.1080/09672559.2014.913888.

Soames, Scott

2014 'Cognitive Propositions'. In *New Thinking about Propositions*. Ed. by Jeffrey C. King, Scott Soames and Jeff Speaks. Oxford University Press, pp. 91–124. DOI: 10.1093/acprof:oso/9780199693764.003.0006.

Soldati, Gianfranco

1998 'The Epistemological Basis of Subjectivity'. Habilitation.

2012 'Direct Realism and Immediate Justification'. *Proceedings of the Aristotelian Society*. 112, pp. 29–44. DOI: 10.1111/j.1467-9264.2012.00324.x.

2013 'Prospects of a Deflationary Theory of Self-Knowledge'. *Studia Philosophica*. 72, pp. 169–187.

2016 'Inferences in the First Person'. *Phenomenology and Mind*. 10, pp. 156–166. DOI: 10.13128/Phe_Mi-20098.

Stalnaker, Robert C.

1999 'Indexical Belief'. In *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford University Press, pp. 130–149.

Strawson, Peter Frederick

1959 *Individuals*. London: Routledge.

Thompson, Evan

2007 *Mind in Life*. Harvard University Press.

Tye, Michael

1995 *Ten Problems of Consciousness*. MIT Press.

Wittgenstein, Ludwig

1922 *Tractatus Logico-Philosophicus*. Trans. by Charles Kay Ogden. Kegan Paul, Trench, Trubner & Co.

Wittgenstein, Ludwig

1953 *Philosophical Investigations*. Ed. by G.E.M. Anscombe and Rush Rheese. Trans. by G.E.M. Anscombe. Blackwell Publishing.

1958 *The Blue and Brown Books*. Oxford: Basil Blackwell.

1961 *Tractatus Logico-Philosophicus*. Trans. by David Pears and Brian McGuinness. Routledge.

Wright, Crispin

1998 'Self-Knowledge: The Wittgensteinian Legacy'. In Wright et al. (1998), pp. 13–45. DOI: 10.1093/0199241406.003.0002.

2001 *Rails to Infinity*. Harvard University Press.

Wright, Crispin, Barry C. Smith and Cynthia Macdonald

1998 (eds.), *Knowing Our Own Minds*. Clarendon Press. DOI: 10.1093/0199241406.001.0001.

Wüstholz, Florian L.

2013 'Selbstbewusstsein bei Tieren: begriffliche und methodologische Probleme'. *Studia Philosophica*. 72, pp. 87–101.

2015 'Self-Consciousness in Animals: Advantages and Problems of a Multipronged Approach'. *Kriterion*. 29, 1, pp. 1–17.

2018 'De Se Beliefs, Self-Ascription, and Primitiveness'. *Disputatio*. 9, 46, pp. 401–422. DOI: 10.1515/disp-2017-0012.

INDEX

COLOPHON

This book was typeset in 10pt *Adobe Caslon Pro*, designed by Carol Twombly, for the body text and Herman Zapf's *Euler* for math and fomulæ in LaTeX. The style was inspired by Robert Bringhurst's seminal book on typography *The Elements of Typographic Style* and adapted and developed for LaTeX by André Miede. Several small adjustments to the layout were made by the author.

Adobe Caslon Pro originates in the work of William Caslon I (1692–1766) who worked as an engraver of punches for typefaces based on seventeenth-century Dutch old style designs. Caslon's types became very popular throughout Europe and the American colonies and the typeface now known as 'Caslon' was used for the first printings of the American Declaration of Independence and the Constitution. Carol Twombly studied pages printed by William Caslon between 1734 and 1770 to arrive at her revival which was published for digital use in 1990. It added many features such as small caps, old style figures, ligatures and ornaments.

The drawing on the titlepage is *Narcissus* by Michelangelo Merisi da Caravaggio from 1599. It shows the mythical Narcissus leaning over water, intently gazing at his own reflection while everything around him is engulfed in darkness, emphasising the engrossing occupation with himself. Caravaggio (1571–1610) was a defining figure for Baroque painting and was known for his highly realistic depiction of the physical and emotional human state. The work, including the faithful photographic reproduction, is a public domain work of art.