

Self in Mind  
A Pluralist Account of Self-Consciousness



Raphaël Millière  
Magdalen College  
University of Oxford

A thesis submitted for the degree of  
*Doctor of Philosophy*

Michaelmas 2019

## Acknowledgements

This work was supported by the Ertegun Graduate Scholarship Programme in the Humanities, the Royal Institute of Philosophy, the Fondation des Treilles, the University of Oxford's Vice Chancellors' Fund, and Magdalen College.

Many people have supported me professionally and personally during the writing of this thesis. My greatest debt of gratitude is owed to my primary supervisor, Martin Davies, whose continued guidance, support, and generosity have been invaluable from the beginning to the end. I have learned a great deal from his detailed and insightful comments and from our many conversations, and I consider myself lucky to have benefited from his wisdom and mentorship. I would also like to express my sincere gratitude to my secondary supervisors: Dan Zahavi, who generously welcomed me at the Center for Subjectivity Research at the University of Copenhagen, who gracefully accepted to discuss matters of philosophical agreement and disagreement with an equally open mind, and whose work inspired me to start thinking about the topic of self-consciousness many years ago as an undergraduate at the École Normale Supérieure; and Philipp Koralus, who provided me with very valuable philosophical and professional advice along the way, and who offered me an opportunity to be a guest lecturer and assessor for the Philosophical Foundations of Psychology module of the University of Oxford's MSc in Psychological Research.

Special thanks are also due to Barry C. Smith, for welcoming me with open arms as a non-stipendiary fellow at the Institute of Philosophy for a year, and for introducing me to the wonderful world of public engagement at the crossroads between philosophy, art and science; and to Thomas Metzinger, for encouraging me to work on less well-trodden topics such as drug-induced and meditation-induced disruptions of self-consciousness, and for his enthusiastic collaboration on the *Selfless Minds* project. I am also very grateful to Bénédicte Veillet and Frédérique de Vignemont for giving me insightful feedback on distant ancestors of chapters 3 and 6, respectively.

I have learned from conversations with many other excellent philosophers and scientists during the writing of this thesis, including Adrian Alsmith, Aviva Berkovich-Ohana, Ophelia Deroy, Chris Frith, Philip Gerrans, Marie Guillot, Uriah Kriegel, Bigna Lenggenhager, Chris Letheby, Jakub Limanowski, Alisa Mandrigin, Salvador Mascarenhas, James Nguyen, Martine Nida-Rümelin, Claire Petitmengin, Miguel Ángel Sebastián, Andrea Serino, Anil Seth, Nick Shea, Enzo Tagliazucchi, and Manos Tsakiris. I

am also grateful to my collaborators and friends from the Centre for Psychedelic Research at Imperial College London, Robin Carhart-Harris, Chris Timmermann, Leor Roseman and Hannes Kettner; and to my colleagues and friends from the ALIUS research group, especially Martin Fortier and Guillaume Dumas.

Writing a doctoral dissertation can be a lonely endeavour; but I have been very fortunate to be surrounded by many people who enriched my life and supported me during challenging times. Thank you Thibaut, Stefano, Julien, Benjamin, Susanna, Dionysios, and Mirte for being such wonderful friends.

I owe my deep gratitude to my family, without whom I would not be where I am: to my parents Jehanne and Stéphane, to my brother Joseph, and to my grand-mother Jacqueline. Thank you for your unconditional love and support.

Finally, there is the one who has shared every moment of this journey with me, and without whom it would have been much duller. Thank you, Ella, for years of love, joy, and laughter.

# Abstract

## *Self in Mind: A Pluralist Account of Self-Consciousness*

74,980 words

This thesis investigates the relationship between consciousness and self-consciousness. I consider two broad claims about this relationship: a *constitutive claim*, according to which *all* conscious experiences constitutively involve self-consciousness; and a *typicalist claim*, according to which *ordinary* conscious experiences contingently involve self-consciousness. Both of these claims call for elucidation of the relevant notions of consciousness and self-consciousness.

In the first part of the thesis ('The Myth of Constitutive Self-Consciousness'), I critically examine the *constitutive claim*. I start by offering an elucidatory account of *consciousness*, and outlining a number of foundational claims that plausibly follow from it. I subsequently distinguish between two concepts of *self-consciousness*: *consciousness of one's experience*, and *consciousness of oneself (as oneself)*. Each of these concepts yields a distinct variant of the constitutive claim. In turn, each resulting variant of the constitutive claim can be interpreted in two ways: on a 'minimal' or deflationary reading, they fall within the scope of foundational claims about consciousness, while on a 'strong' or inflationary reading, they point to determinate aspects of phenomenology that are not acknowledged by the foundational claims as being aspects of all conscious mental states. I argue that the deflationary readings of either variant of the constitutive claim are plausible and illuminating, but would ideally be formulated without using a term as polysemous as 'self-consciousness'; by contrast, the inflationary readings of either variant are not adequately supported.

In the second part of the thesis ('Self-Consciousness in the Real World'), I focus on the second concept of self-consciousness, or *consciousness of oneself as oneself*. Drawing upon empirical evidence, I defend a *pluralist account* of self-consciousness so construed, according to which there are several ways in which one can be conscious of oneself as oneself – through conscious thoughts, bodily experiences and perceptual experiences – that make distinct determinate contributions to one's phenomenology. This pluralist account provides us with the resources to vindicate the *typicalist claim* according to which consciousness of oneself as oneself – a *sense of self* – is pervasive in ordinary conscious experiences, as a matter of contingent empirical fact. It also provides us with the resources to assess the possibility that a subject might be conscious without being conscious of herself as herself in any way.

# Contents

<b>Introduction: Consciousness and Self-Consciousness</b>	<b>1</b>
<b>I THE MYTH OF CONSTITUTIVE SELF-CONSCIOUSNESS</b>	
<b>1 Elucidating Consciousness</b>	<b>6</b>
1.1 Conscious creatures and conscious mental states . . . . .	7
1.2 Access consciousness and phenomenal consciousness . . . . .	8
1.3 Characterising consciousness by example . . . . .	10
1.4 The Nagelian dictum . . . . .	12
1.4.1 The technical and comparative accounts of the Nagelian dictum . . . . .	14
1.4.2 The affective account of the Nagelian dictum . . . . .	16
1.5 The phenomenal character of consciousness . . . . .	24
1.6 Foundational claims about consciousness . . . . .	27
1.6.1 Claims about the (self-)ascription of conscious mental states . . . . .	27
1.6.2 Claims about the subject of experience . . . . .	34
<b>2 Two Concepts of Self-Consciousness</b>	<b>37</b>
2.1 Defining self-consciousness . . . . .	38
2.1.1 Self-consciousness as consciousness <i>of consciousness itself</i> . . . . .	38
2.1.2 Self-consciousness as consciousness <i>of oneself</i> . . . . .	40
2.1.3 A deflationary notion of self-consciousness? . . . . .	41
2.1.4 A note on terminology . . . . .	42
2.2 Two constitutive claims: a first pass . . . . .	43
2.3 Historical overview . . . . .	45
2.3.1 Non-egological self-consciousness . . . . .	45
2.3.2 Egological self-consciousness . . . . .	58
2.4 Conclusion . . . . .	63
<b>3 Debunking the Myth: Non-Egological Self-Consciousness</b>	<b>65</b>
3.1 From subjectivity to consciousness <i>of one's experience</i> . . . . .	66
3.2 Zahavi's minimal constitutive claim . . . . .	76
3.3 Kriegel's substantive constitutive claim . . . . .	81

3.4 Conclusion . . . . .	91
<b>4 Debunking the Myth: Egological Self-Consciousness</b>	<b>92</b>
4.1 From subjectivity to consciousness of <i>oneself</i> . . . . .	93
4.2 Minimal constitutive claims . . . . .	96
4.3 Substantive constitutive claims . . . . .	108
4.4 Non-constitutive substantive claims . . . . .	112
<b>II SELF-CONSCIOUSNESS IN THE REAL WORLD</b>	
<b>5 Cognitive Self-Consciousness</b>	<b>120</b>
5.1 Thinking about oneself as oneself . . . . .	122
5.2 <i>De se</i> thought and self-consciousness . . . . .	125
5.3 Is cognitive self-consciousness ubiquitous? . . . . .	136
5.4 Cognitive self-consciousness in ordinary experience . . . . .	140
5.5 Conclusion . . . . .	143
<b>6 Bodily Self-Consciousness</b>	<b>145</b>
6.1 Non-conceptual <i>de se</i> content . . . . .	146
6.2 The sense of bodily ownership . . . . .	151
6.3 The argument from somatoparaphrenia . . . . .	153
6.4 The argument from drug-induced disownership . . . . .	163
6.5 From body to self . . . . .	167
6.6 Bodily self-consciousness in ordinary experience . . . . .	172
6.7 Conclusion . . . . .	174
<b>7 Spatial Self-Consciousness</b>	<b>175</b>
7.1 Perspectival spatial content . . . . .	177
7.2 <i>de hinc</i> and <i>de se</i> content . . . . .	182
7.3 The argument from film-watching . . . . .	186
7.4 The argument from virtual reality . . . . .	189
7.5 The agency-involving account revisited . . . . .	197
7.6 Spatial self-consciousness in ordinary experience . . . . .	201
7.7 Conclusion . . . . .	202
<b>Conclusion: The Significance of Self-Consciousness</b>	<b>204</b>
<b>References</b>	<b>213</b>

# Introduction:

## Consciousness and Self-Consciousness

Like many animals, we are conscious of the world around us. We enjoy a wealth of experiences that generally provide us with useful information about our environment. Most of the experiences we have concern external objects, such as the desk I see in front of me as I write these words, or the bell of Magdalen Tower I hear ringing in the background. But there is more to what we can experience, for we also have the capacity to be *self*-conscious: we can be conscious of ourselves and of our own experiences.

Self-consciousness is at once familiar and perplexing. It is such a central feature of our mental lives that many philosophers have deemed it pervasive in ordinary conscious experience.<sup>1</sup> There is some intuitive appeal to the idea that in the wakeful state, we are not simply conscious of the world around us, but also of ourselves and of our inner mental life. However, this is a slippery intuition that may vanish as soon as we try to probe our experience. David Chalmers eloquently articulates this observation:

One sometimes feels that there is something to conscious experience that transcends all [other] specific elements: a kind of background hum, for instance, that is somehow fundamental to consciousness and that is there even when the other components are not. This phenomenology of self is so deep and intangible that it sometimes seems illusory, consisting in nothing over and above [perceptions, sensations, thoughts or emotions]. Still, there seems to be something to the phenomenology of self, even if it is very hard to pin down.

Chalmers (1996, p. 10)

A number of authors endorse a stronger claim still, according to which self-consciousness would be *constitutive of* consciousness.<sup>2</sup> On this view, all conscious experiences necessarily

<sup>1</sup>See, for example, Block (1995, p. 235), Kriegel (2009, p. 177), Strawson (2011b, p. 8), and Guillot (2017, p. 46), among others.

<sup>2</sup>See, for example, Damasio (1999, p. 19), S. Gallagher (2010, p. 567), Zahavi (2014, p. 17), Duncan (2019, p. 301), Friston (2018, p. 1), among others.

and constitutively involve a form of *self*-consciousness; in being conscious at all, one is *ipso facto self*-conscious. Call this the *constitutive claim*. Intuitions about the constitutive claim are starkly divided. Some find it obviously true (e.g., Zahavi 2014), while others find it clearly false (e.g., Metzinger 2003). But intuitions can only get us so far; what we need to arbitrate this disagreement is a philosophical elucidatory account of consciousness and self-consciousness.

The broad aim of this thesis is to illuminate the relationship between consciousness and self-consciousness. Since both notions are notoriously polysemous, this aim can only be achieved through careful disambiguation and conceptual analysis. As a result, there might be many ways in which one can adequately characterise the relationship between consciousness and self-consciousness – depending on what one means by ‘consciousness’, and, crucially, by ‘self-consciousness’.

The thesis is divided in two parts that differ in a number of respects. Part I is, as it were, the *pars destruens*: it critically examines the constitutive claim, or, as I provocatively call it, the ‘myth’ of constitutive self-consciousness. It is written in close dialogue with philosophers past and present who have defended the constitutive claim, and the resulting discussion remains mostly at the level of *a priori* philosophical theorising. Part II is, by contrast, the *pars construens*: it vindicates the weaker but still substantive claim that self-consciousness, although not constitutive of consciousness, is nonetheless a component of ordinary conscious experience – in a sense to be further elucidated. Compared to Part I, it is less steeped in the history of philosophy, but engages much more closely with relevant empirical work from psychology and neuroscience.

The more specific aim of each chapter can be summarised as follows:

- **Chapter 1** provides a philosophical elucidatory account of consciousness that draws upon Thomas Nagel’s influential proposal, and outlines a number of foundational claims that plausibly follow from this account.
- **Chapter 2** distinguishes two broad concepts of self-consciousness, namely consciousness of *consciousness itself*, and consciousness of *oneself*; each of these concepts yields

a distinct variant of the constitutive claim, whose ancestry in the history of philosophy is surveyed.

- **Chapter 3** examines the first variant of the constitutive claim, namely the claim that consciousness constitutively involves consciousness *of consciousness itself* (or consciousness *of one's experience*).
- **Chapter 4** examines the second variant of the constitutive claim, namely the claim that consciousness constitutively involves consciousness *of oneself*.
- **Chapter 5** shifts the focus of the discussion to the weaker claim that ordinary experience involves phenomenal self-consciousness, defined as consciousness *of oneself as oneself*; it examines the most obvious way in which one can be conscious *of oneself as oneself*, namely by engaging in conscious thinking *about oneself as oneself* ('cognitive self-consciousness').
- **Chapter 6** considers whether one can be phenomenally self-conscious without consciously thinking about oneself as oneself. Specifically, it is argued that in ordinary circumstances, feeling a sensation in a body part is *ipso facto* being phenomenally self-conscious, in so far as one is non-conceptually conscious of oneself as the bodily subject whose body part one feels the sensation in or on ('bodily self-consciousness').
- **Chapter 7** argues that there is a third form of phenomenal self-consciousness that is even more pervasive in ordinary experience: in ordinary circumstances, undergoing perceptual experiences with perspectival spatial content is *ipso facto* being phenomenally self-conscious, in so far as one is non-conceptually conscious of oneself as being located at the origin of the experience's spatial perspective ('spatial self-consciousness').

The upshot of this analysis is that while phenomenal self-consciousness – consciousness *of oneself as oneself* – is not constitutive of all conscious experiences, it is nonetheless pervasive in ordinary conscious experiences, in so far as spatial self-consciousness (and, to a lesser extent, bodily self-consciousness) is. In conclusion, I offer some reflections about three outstanding questions:

- (i) Are there *other* forms of phenomenal self-consciousness, beyond *cognitive*, *bodily* and *spatial* self-consciousness?
- (ii) Are there states of consciousness from which *all* forms of phenomenal self-consciousness are missing?
- (iii) What is the significance of phenomenal self-consciousness – its epistemic significance for rational creatures, and more generally its biological significance for conscious organisms?

I

THE MYTH OF CONSTITUTIVE  
SELF-CONSCIOUSNESS

# 1

## Elucidating Consciousness

Many fall into the trap of confusing consciousness with self-consciousness...

Sutherland (1989, p. 90)

In order to investigate the nature of the relationship between consciousness and self-consciousness, it is crucial to gain some clarity on the meaning of each of these terms. The term ‘consciousness’ is notoriously difficult to define, and has multiple uses that we may not be able to reconcile into a single coherent notion. In this chapter, I will offer an elucidatory account of the notion of consciousness that is relevant to various versions of the claim that consciousness constitutively involves self-consciousness. This account shall provide the necessary background for the discussion of the constitutive claim in subsequent chapters.

As a starting point to unravel the uses of the term ‘consciousness’, it is useful to consider the main ways in which the adjective ‘conscious’ can be meaningfully predicated. In §1.1, I distinguish between consciousness as a predicate of creatures, and consciousness as a predicate of mental states. In §1.2, I subsequently distinguish between the technical notion of *access consciousness*, and the non-technical notion of *phenomenal consciousness*, on which the rest of this chapter focuses. In §1.3, I consider attempts to characterise the notion of phenomenal consciousness *by example*, and emphasise the limitations of such an approach. In §1.4, I examine instead Nagel’s characterisation of phenomenal consciousness *by elucidation*, according to which a mental state of a subject is a *conscious* mental state of that subject if and only if there is *something it is like* for that subject to be in that mental state. After considering three interpretations of Nagel’s elucidatory account, I endorse the *affective* interpretation, according to which a mental state of a subject is a

*conscious* mental state of that subject if and only if it is *constitutively* such that it affects that subject in an experiential way, or, equivalently, it *constitutively* contributes to that subject's overall phenomenology. In §1.5, I consider some implications of this elucidatory account for the distinction between there being *something*, rather than *nothing*, that it is like to be in a conscious mental state, and *what* it is like to be in conscious mental state. Finally, in §1.6, I discuss several foundational claims about consciousness that plausibly follow from the Nagelian elucidatory account of consciousness.

### 1.1 Conscious creatures and conscious mental states

The adjective 'conscious' is often predicated of *creatures*, where this notion refers not only to biological organisms such as humans and cephalopods, but more generally to information-processing systems broadly construed, including artificial systems.<sup>1</sup> For example, one might say of a horse, Bucephalus, that he is *conscious* – rather than *non-conscious*.

There is an ambiguity with the use of 'conscious' as a predicate of creatures: it can refer either to a *dispositional* property or to a *categorical* property of creatures. Thus, one might say that Bucephalus is conscious *at a specific time*, for example at a time when Bucephalus is trotting in a field, to mean that Bucephalus instantiates the categorical property of being conscious at that time. Alternatively, one might say that Bucephalus is a conscious creature in general (namely, in the dispositional sense), to mean that Bucephalus has the *general capacity* to be conscious (in the categorical sense), although he might not actually be conscious (in the categorical sense) at the time of utterance – for example, if he happens to be fully anaesthetised with ketamine at that time. As this example shows, the dispositional use of 'conscious' as a predicate of creatures appears to be grounded in the categorical use of 'conscious' as a predicate of creatures; consequently, we can leave aside the dispositional use for our purposes.

The adjective 'conscious' can be predicated not only of creatures, but also of the *mental*

<sup>1</sup>Rosenthal (1986) refers to this use of 'conscious' as 'creature-consciousness'.

*states* (or *mental events*) of creatures.<sup>2</sup> For example, one might say that a human creature, Mary, has a *conscious* desire to eat ice cream; meaning that Mary's desire to eat ice cream is a *conscious* desire of Mary, as opposed to a *non-conscious* desire of Mary. This is *prima facie* equivalent to the adverbial construction 'Mary *consciously* desires to eat ice cream'.

Importantly, the two uses of the adjective 'conscious' as a predicate of creatures and as a predicate of mental states are not independent: it is generally agreed that a creature is *conscious* at time *t* if and only if it is in a *conscious* mental state at *t*. For example, if Mary has a desire to eat ice cream at *t*, and this desire is a *conscious* desire of Mary at *t*, then Mary is conscious at *t*. Conversely, if Mary is conscious at *t*, then she is in a *conscious* mental state at *t*; for example, she has the *conscious* desire to eat ice cream at *t*.

## 1.2 Access consciousness and phenomenal consciousness

These preliminary remarks pertain to the patterns of predication of the adjective 'conscious' rather than to its meaning. What does it mean for a creature to be *conscious* rather than *non-conscious*? And what does it mean for a mental state of a creature to be a *conscious* mental state of that creature, as opposed to a *non-conscious* mental state of that creature? There are different ways to answer these question, which result in distinct notions of consciousness.

One answer consists in defining consciousness as a technical term, and distinguishing *conscious* and *non-conscious* mental states of a creature with respect to their functional properties. Specifically, one might say that a mental state *M* of a creature *C* is a *conscious* mental state of *C* if and only if *M*'s content is available for use in *C*'s reasoning and in the rational guidance of *C*'s speech and action. The resulting notion is what Block (1995) calls *access consciousness* or *A-consciousness*.

Note that access consciousness is not defined as a property of creatures, but as a property of mental states. Nonetheless, one could define a corresponding notion of *A-consciousness\** for creatures as follows: a creature *C* is *A-conscious\** if and only if *C* is in a mental state *M* such that *M*'s content is available for use in *C*'s reasoning and in the rational guidance of

<sup>2</sup>Rosenthal (1986) refers to the use of 'conscious' as a predicate of mental states as 'state-consciousness'. Note that I will subsequently use 'mental states' as shorthand for 'mental states or events' for the sake of convenience, although I do not wish to commit to a specific ontology of mental phenomena.

C's speech and action. Thus, if Mary is able to report on her visual perception of a tree (for example by saying 'I see a tree'), this visual perception is an *A-conscious* mental state of Mary (and Mary herself is an *A-conscious\** creature).

However, the definition of access consciousness does not seem to match the folk-psychological concept of consciousness commonly expressed by the adjective 'conscious'. When one says that Bucephalus is *conscious*, one does not normally mean that Bucephalus has a mental state whose content is available for use in reasoning and for rationally guiding speech and action. Likewise, when one says that Mary has a *conscious* desire for ice cream, one does not normally mean that the content of Mary's desire for ice cream is available for use in Mary's reasoning and for the rational guidance of Mary's speech and action. It might well be the case that all the mental states of Bucephalus and Mary (and any other creature) that would commonly be said to be 'conscious' do instantiate the functional properties specified in the definition of access consciousness.<sup>3</sup> Nonetheless, when one says that a creature or a mental state is conscious in a pre-theoretical context, one presumably has in mind a concept that is less theoretically sophisticated than the technical concept of access consciousness.

By contrast with *access consciousness*, the non-technical or folk-psychological concept of consciousness is often called *phenomenal consciousness* (Block 1995). Phenomenal consciousness is the concept that one usually latches onto when one says that a creature is *conscious*, or that the mental state of a creature is a *conscious* mental state of that creature. At a first pass, one might say that "phenomenal consciousness is experience" (Block 1995, p. 228). But substituting one undefined term for another is hardly helpful to capture its meaning, if one does not have a prior understanding of what the term is supposed to refer to. Moreover, as it has often been remarked, it seems impossible to define the notion of phenomenal consciousness analytically, as one can define a triangle as a plane figure with three edges and three vertices. This is because the meaning of the relevant notion presumably cannot be grasped *in abstracto* in terms of something more fundamental (by

<sup>3</sup>Of course, it is dubious that a horse like Bucephalus could verbally report on any of his mental states; but the availability of a mental state's content for verbal report need not be a necessary condition of access consciousness.

contrast with the technical concept of access consciousness). One can understand what a triangle is without ever seeing a triangle; but one cannot understand what phenomenal consciousness is without having been in a phenomenally conscious mental state.

I will now explore two strategies to characterise phenomenal consciousness, by *example* and by *elucidation* respectively. As we shall see in the following chapters, philosophical discussions regarding the relationship between consciousness and self-consciousness typically construe the former as *phenomenal consciousness* rather than *access consciousness*. For the sake of simplicity, I will henceforth refer to phenomenal consciousness – the folk-psychological concept distinct from the technical concept of *access consciousness* – as ‘consciousness’ *simpliciter*, unless otherwise stated.

### 1.3 Characterising consciousness by example

A common strategy to characterise consciousness as a property of creatures or as a property of mental states is to provide *examples* of conscious creatures and conscious mental states. Call this the *illustrative strategy*. Assuming that the relevant notion of consciousness is an obvious folk-psychological concept, the hope is that providing some examples in which this concept applies (and others in which it does not) will be enough to pin down its meaning. In other words, the illustrative strategy relies on the readers’ ability to latch onto the concept of consciousness by generalising from a set of examples.

We can illustrate the meaning of consciousness as a predicate of creatures by discussing cases in which some creature is conscious, and cases in which the same creature is not conscious.<sup>4</sup> Consider again our horse, Bucephalus. When Bucephalus is fully anaesthetised with ketamine, he is not conscious; as he wakes up from general anaesthesia, however, Bucephalus becomes conscious. Likewise, when Mary is in a deep, dreamless sleep, she is not conscious; when she wakes up in the morning, by contrast, she is conscious. Generally, when creatures with the capacity for consciousness are awake and sober – rather than asleep, anaesthetised or in a coma – they are conscious.

<sup>4</sup>See, for example, Searle (1992, p. 83).

Similarly, we can illustrate the meaning of consciousness as a predicate of mental states by providing positive examples of conscious mental states and negative examples of non-conscious mental states.<sup>5</sup> A list of mental states that are typically *conscious* mental states might include *tasting a sweet orange, seeing a yellow lemon, imagining a sunset, hearing a C sharp, feeling a sharp pain, feeling cold, or feeling angry*; it may also include more contentious cases such as *thinking about the future, remembering a joke, calculating a multiplication, desiring ice cream or feeling depressed*. In turn, a list of mental states that are typically *non-conscious* mental states might include *knowing the syntactical rules of English, subliminally seeing an advertisement for Coca-Cola, decoding vestibular signals to balance one's body, being disposed to answer 'four' to the question 'two plus two equals?', or perhaps having a repressed desire to quit one's job*.

The illustrative strategy relies on the assumption that these examples will allow the reader to grasp the relevant folk-psychological concept of consciousness, by abstracting away from their differences to capture their generic commonality – namely, a feature that all the positive examples share and that all the negative examples lack. Importantly, the target concept should be an obvious one, rather than a sophisticated philosophical concept such as *access consciousness*.

The success of the illustrative strategy as a way of characterising the concept of consciousness is debatable. Some think it is the best strategy, because it allows us to characterise consciousness as innocently as possible, without building controversial substantive assumptions into the concept (e.g., Schwitzgebel 2016). Others think this strategy is hopeless, because they think it is unclear what – if anything – all positive examples might have in common that all negative examples lack (e.g., Frankish 2016). Part of the problem is that the illustrative strategy does not provide a *positive* characterisation of the concept of consciousness (as a property of creatures or as a property of mental states) – it does not directly elucidate what consciousness *is*. For this reason, the illustrative strategy is often supplemented with a distinct strategy, which consists in *elucidating* the concept of

<sup>5</sup>See, for example, Chalmers (1996, p. 4).

consciousness by providing intuitive paraphrases of what it means for a creature or for a mental state to be a *conscious* creature or a *conscious* mental state.

#### 1.4 The Nagelian dictum

The most influential attempt at characterising the notion of consciousness by elucidation is due to Thomas Nagel's seminal article "What is it like to be a bat?" (Nagel 1974). Nagel elucidates the notion in the following way:

[T]he fact that an organism has conscious experience *at all* means, basically, that there is something it is like to *be* that organism... [F]undamentally an organism has conscious mental states if and only if there is something that it is like to *be* that organism – something it is like *for* the organism.

Nagel (1974, p. 436)

There is a lot to unpack in this passage. At a first pass, we can condense it into the following Nagelian dictum:

(N) An organism has *conscious* mental states if and only if there is something it is like for that organism to be that organism.

This dictum calls for three observations. Firstly, it refers to organisms rather than creatures. However, this does not seem to reflect substantial assumptions on Nagel's part concerning the restriction of his account to *biological organisms*, as opposed to *creatures* in general (including artificial systems). Thus, it is relatively innocuous to substitute 'creature' for 'organism' in the dictum.

Secondly, while the term 'conscious' is predicated of *mental states* in (N), the claim is not presented as an elucidation of what it is for a mental state to be conscious rather than non-conscious, but is rather focused on what it is for an organism to be in conscious mental states. With minor modifications, however, the dictum can be adapted to provide an elucidatory account of consciousness either as a property of creatures, or as a property of mental states.

Thirdly, the Nagelian dictum is slightly ambiguous with respect to the aforementioned distinction between categorical and dispositional uses of 'conscious', in part because it refers to *conscious mental states* in plural form (rather than to *a specific* conscious mental

state), and in part because it is not restricted to a specific time at which a creature is in a conscious mental state. Thus, one could perhaps read (N) as claiming that an organism has *the capacity to have* conscious mental states in general if and only if there is something it is like to be that organism. This dispositional interpretation also comes through when Nagel writes “the essence of the belief that bats have experience is that there is something that it is like to be a bat” (Nagel 1974, p. 438). Presumably, Nagel does not mean that bats *always* have conscious mental states, or that there is *always* something it is like to be a bat (including, for example, when bats are in deep, dreamless sleep). Rather, he means that bats have the capacity to have conscious mental states, and often do have such states. However, the Nagelian dictum (N) can be easily modified to remove this ambiguity.

Taking these three points into consideration, we can formulate two claims in the vicinity of (N) that capture the spirit of Nagel’s elucidatory account of consciousness while avoiding ambiguities. The following claim offers a Nagelian account of consciousness as a categorical property of *creatures*:

(N<sub>CR</sub>) A creature C is *conscious* at time *t* if and only if there is something it is like for C to be C at *t*.

As we have seen, the use of ‘conscious’ as a predicate of mental states is related to the use of ‘conscious’ as a predicate of creatures, in the following sense: a creature is *conscious* at time *t* if and only if the creature is in a *conscious* mental state at *t*. Accordingly, we can formulate a distinct claim offering a Nagelian account of consciousness as a categorical property of *mental states*:

(N<sub>ST</sub>) A mental state M of a creature C is a *conscious* mental state of C at time *t* if and only if there is something it is like for C to be in M at *t*.

The phrase ‘something it is like’ (henceforth, the *SIL phrase*) plays a central role in all three of these claims. In his article, Nagel also employs the related phrase ‘what it is like’ (henceforth, the *WIL phrase*), for example when he writes “I want to know what it is like for

a *bat* to be a bat” (Nagel 1974, p. 439).<sup>6</sup> There is a great deal of disagreement about how to interpret the WIL and SIL phrases, and whether or not they are helpful at all to elucidate the meaning of consciousness – as a property of creatures and as a property of mental states.

### 1.4.1 The technical and comparative accounts of the Nagelian dictum

A first line of criticism of the use of the WIL and SIL phrases to elucidate consciousness is that such use does not reflect the ordinary meaning of these phrases, but is rather intended to have a technical meaning that is ultimately not very clear (e.g., Lewis 1995, p. 140). However, Nagel introduces neither of these phrases as a technical expression, and one can find many instances of the WIL phrase to characterise conscious experiences outside of philosophical works. For example, in the novel *It Is Never Too Late to Mend* by Charles Reade (1856), a chaplain appalled by the use of a painful restraining jacket as punishment in a prison decides to try the device on himself to know what inmates have to endure. He subsequently has the following exchange with a prison guard who has never had this experience himself, and only knows (by observing inmates suffer) that it must be ‘uncomfortable’:

PRISON GUARD: What is it like, sir? haw! haw!

CHAPLAIN: It is, as you described it, *uncomfortable*; but the knowledge I have gained in it is invaluable.

Reade (1856, p. 164)

This passage significantly predates any philosophical use of the WIL phrase (or the SIL phrase, for that matter) as an elucidation of consciousness.<sup>7</sup> Yet it seems perfectly natural

<sup>6</sup>Nagel was not the first philosopher to use these phrases in relation to consciousness, although he was arguably the first to explicitly offer an elucidation of consciousness by using these phrases systematically. Earlier philosophical uses of the WIL phrase include Russell (1926), Wittgenstein (1947/1980, §91), B. A. Farrell (1950, p. 181), and Sprigge (1971, p. 167).

<sup>7</sup>For later examples of non-philosophical uses of the WIL phrase in reference to conscious experience, see J. Farrell (2016, p. 60) who quotes Anton Chekhov’s play *The Seagull* (1865) in Marian Fell’s English translation: “What is it like to be famous? What sensations does it give you?”; and Stoljar (2016, p. 1163) who quotes the song *To Love Somebody* by the Bee Gees (1967): “You don’t know what it’s like to love somebody”. Note that the original Russian text of the passage from Chekhov (“Как чувствуется известность? Как вы ощущаете то, что вы известны?”) translates more literally to “How does fame feel? How do you feel being famous?” (I am grateful to Julius Kochan for this literal translation). Still, it is interesting to note that Marian Fell’s translation, which was the first English translation of *The Seagull* to be published in the United States in 1912, uses the WIL phrase as the most natural rendition of the Russian equivalent of ‘how does *x* feel’.

for the prison guard to ask *what it is like* to put on the restraining jacket. Furthermore, if one can naturally ask *what it is like* to put on the restraining jacket, it should also be natural to say that there is *something it is like* to put on the restraining jacket. Thus, it is dubious that the WIL phrase, and consequently the SIL phrase, should be intended in a special, technical sense when they are used to elucidate the meaning of consciousness, as in (N<sub>CR</sub>) and (N<sub>ST</sub>).

A second line of criticism of the WIL and SIL phrases consists in claiming that if these phrases are not technical terms, then they must express judgements of similarity, because this is what statements of the form ‘*x is like y*’ normally express in English. For example, one might ask “What is the colour of lapis lazuli like?” to prompt a comparative statement, such as “the colour of lapis lazuli is like the colour of a mountain lake”. Similarly, one might ask “Is there anything that the colour of lapis lazuli is like?” to ask whether the colour of lapis lazuli is similar to any other colour; and if it is impossible to produce such judgement of similarity, one might reply “there is nothing that it is like” or “it is hard to say what it is like”.

Note that Nagel himself rejected the idea that the WIL and SIL phrases express comparative judgements:

[T]he analogical form of the English expression ‘what it is *like*’ is misleading. It does not mean ‘what (in our experience) it *resembles*’, but rather ‘how it is for the subject himself’.

Nagel (1974, p. 440, fn. 6)

That the WIL and SIL phrases do not express comparative judgements is clear when we consider statements such as “there is nothing it is like to be stone” (Crane 1999, p. 548). This statement does not mean that there is nothing that *resembles* being a stone; being a boulder, a brick or a grain of sand resemble being a stone in some respects. Thus, one could meaningfully say “there is something *that is like being a stone*” with a comparative reading in mind. But one could not equivalently say “there is something *that it is like to be a stone*” with such a reading in mind, let alone “there is something *that it is like to be a stone* – something that it is like *for the stone*” (paraphrasing the Nagelian dictum). Thus, the comparative reading of the Nagelian dictum does not hold.

### 1.4.2 The affective account of the Nagelian dictum

Some authors suggest that if the WIL and SIL phrases are neither technical terms nor comparative expressions, their meaning is ultimately obscure.<sup>8</sup> However, there is a non-technical use of these phrases that is distinct from their use in judgements of similarity.

Recall that Nagel, in his rejection of the comparative account, stresses that the WIL phrase refers to ‘how it is *for the subject himself*’ (my emphasis). Likewise, in passages that formulate the Nagelian dictum, he puts a specific emphasis on there being “something it is like *for the organism*” (Nagel 1974, p. 436, emphasis in the original). This emphasis is significant to grasp the meaning of the Nagelian dictum. To see this, it is useful to unpack the grammar of sentences containing the WIL phrase or the SIL phrase as a way to characterise consciousness, and particularly the role of the preposition ‘for’.

Following Stoljar (2016), we can note that there are two potentially implicit argument places in a sentence such as:

- (a) There is something it is like to have a toothache.<sup>9</sup>

There is a first argument place in (a) for the covert subject of the infinitive verb ‘to have’ in the embedded clause ‘to have a toothache’. For this subject to be overtly mentioned in the sentence, one needs to add a prepositional phrase ‘for NP’, where ‘NP’ is a noun phrase (e.g., a proper name, a pronoun, or a description). For the sake of clarity, I shall indicate this use of ‘for’ to specify the covert subject of a *to*-infinitival construction by ‘for<sup>SUBJ</sup>’, as in:

- (b) There is something it is like for<sup>SUBJ</sup> Mary to have a toothache.

Used in this way, ‘for’ specifies *who* has a toothache. Note that in (b), ‘it’ refers to the embedded clause ‘for<sup>SUBJ</sup> Mary to have a toothache’. Thus, we can substitute this clause for ‘it’ to simplify the syntax of (b), in order to emphasise the subject-specifying role of ‘for’:

- (c) For<sup>SUBJ</sup> Mary to have a toothache is like something.

<sup>8</sup>See for example Hacker (2002) and J. Farrell (2016).

<sup>9</sup>Stoljar (2016)’s example (1a), p. 1163. I shall label examples with lower-case letters.

Furthermore, we can preserve the meaning of (b) and (c) by removing the ‘for’ entirely, as in:

(d) That Mary has a toothache is like something.

And:

(e) Mary’s having a toothache is like something.

This makes it explicit that it is *Mary’s* having a toothache, rather than *anyone else’s*, that is like something.

The second potentially implicit argument place in (a) is for the covert indirect object of the impersonal verbal form ‘it is’ in the main clause ‘there is something it is like’. Again, the indirect object can be overtly specified with a prepositional phrase ‘for NP’, where ‘NP’ is a noun phrase. For the sake of clarity, I shall indicate this use of ‘for’ to specify the indirect object of a verb by ‘for<sup>OBJ</sup>’, as in:

(f) There is something it is like for<sup>OBJ</sup> Mary to have a toothache.

Confusingly, (f) looks exactly like (b) when the syntactical role of ‘for’ is not made explicit. However, while in (b) ‘for’ specifies *who* has a toothache, in (f) ‘for’ specifies *for whom* it is like something (to have a toothache).<sup>10</sup> Again, we can make the structure of (f) more explicit by reformulating it in the following way:

(g) Having a toothache is like something for<sup>OBJ</sup> Mary.

This makes it explicit that it is *for Mary*, rather than *for anyone else*, that having a toothache is like something.

Sticking with the example of Mary, we can now fill both implicit argument places in (a):

<sup>10</sup>Note that this ambiguity is fairly common in the English language. For example, “it is important for us to clean” can mean either “That we clean is important” or “Cleaning is important for us”, depending on whether ‘for’ is interpreted as for<sup>SUBJ</sup> or for<sup>OBJ</sup> respectively. There is a subtle semantic difference between the two interpretations: the former does not explicitly specify *for whom* it is important that we clean, while the latter does not explicitly specify *whose* cleaning is important for us.

(h) There is something it is like for<sup>OBJ</sup> Mary, for<sup>SUBJ</sup> Mary to have a toothache.

Again, note that we can make the structure of this sentence more explicit by getting rid of ‘it’ and ‘for<sup>SUBJ</sup>’:

(i) Mary’s having a toothache is like something for<sup>OBJ</sup> Mary.

This example should already alert us to the fact that the indirect object of the verb ‘is’ (specified by ‘for<sup>OBJ</sup>’) is particularly significant in the Nagelian account of consciousness; however, more work is required to explain how this is significant. The distinction between the two potentially covert argument places in the sentence (a) is not sufficient to guarantee that the sentence will be read in the appropriate way, namely as pertaining to the *conscious experience* of having a toothache. Indeed, we can find plenty of sentences that have nothing to do with consciousness, in which an impersonal verbal construction with an embedded *to*-infinitival clause can be completed with an overt subject introduced by ‘for<sup>SUBJ</sup>’ and an overt indirect object introduced by ‘for<sup>OBJ</sup>’. Consider, for example:

(j) It is unfortunate for<sup>OBJ</sup> the UK, for<sup>SUBJ</sup> the UK to leave the EU.

Clearly, this sentence does not imply that the UK *experiences* its leaving the EU. This is not merely due to the fact that the UK is a country rather than a person, as the following example shows:

(k) It is useful for<sup>OBJ</sup> Mary, for<sup>SUBJ</sup> Mary to know where France is.

Again, this sentence does not imply that Mary *experiences* her knowing where France is – that is, it does not imply that Mary’s knowledge of France’s geographical location is a *conscious* mental state of Mary.

Let us now come back the SIL phrase:

(h) There is something it is like for<sup>OBJ</sup> Mary, for<sup>SUBJ</sup> Mary to have a toothache.

As we have seen, it is very implausible to interpret the SIL phrase as expressing a comparison. We can now understand why: if the SIL phrase in (h) was expressing a comparison, the specification of the indirect object ('for<sup>OBJ</sup> Mary') would make little sense. Although it would be a bit convoluted, one could perhaps understand a sentence such as (e) ("Mary's having a toothache is like something") as meaning that Mary's having a toothache is *similar* to something – e.g., to Mary's having a fever, or to Mary's holding her jaw. But it would seem particularly unnatural to interpret a sentence like (h) as meaning that Mary's having a toothache is similar to something *for Mary*. In sentences like (h), where the indirect object of the main clause is a creature (Mary), and the direct object of the embedded *to*-infinitival clause is a mental state (a toothache), it is natural to interpret 'for<sup>OBJ</sup>' as specifying the *subject of experience*, namely the *psychological subject* whose experience is being talked about. For clarity, I shall indicate this special use of 'for<sup>OBJ</sup>' to specify the subject of experience by 'for<sup>EXP</sup>'.

Note that the WIL and SIL phrases allow the possibility that the subject of experience (i.e., the indirect object of 'is') may be distinct from the subject of the infinitive verb, as in:

- (l) There is something it is like for<sup>EXP</sup> Paul, for<sup>SUBJ</sup> Mary to have a toothache.

What Mary's having a toothache is like for<sup>EXP</sup> Paul might include, for example, hearing Mary groan, seeing Mary rush to the medicine cabinet, or having feelings of sympathy towards Mary. Thus, when both the subject of experience and the subject of the infinitive verb are *explicit*, co-reference is not required: they need not be one and the same subject. However, when both the subject of experience and the subject of the infinitive verb are *implicit* – as in the sentence (a) – or when *only* the subject of experience is explicit – as in the sentences (f) and (g) –, co-reference is required. Thus, "There is something it is like for<sup>EXP</sup> Mary to have a toothache" can only mean: "There is something it is like for<sup>EXP</sup> Mary, for<sup>SUBJ</sup> Mary to have a toothache".

On Stoljar (2016)'s *affective* account of the semantics of sentences containing the SIL phrase or the WIL phrase, a sentence like (a) "means in effect there is *a way* that having a toothache affects you" (p. 1172; my emphasis). Such sentences are context-sensitive and their rather non-specific linguistic meaning allows that, in what Stoljar calls "*stereotypical*

contexts of use” (*ibid.*, p. 1175), a sentence like (a) is used to express the proposition that there is not just *a way* but, more specifically, *an experiential way* that having a toothache affects the subject, as in:

(m) There is *an experiential way* that having a toothache affects Mary.

Stoljar also glosses the stereotypically expressed proposition, “There is some experiential way that *y*’s having a toothache affects *x*”, using the verb ‘feels’: There is some way that *x feels* as a result of *y*’s having a toothache.<sup>11</sup> Thus, (m) can be glossed as follows:

(n) There is some way that Mary *feels* as a result of Mary’s having a toothache.

We can now bring Stoljar’s (2016) account of the logical form and the semantics of sentences containing the SIL phrase or the WIL phrase into contact with the Nagelian account of a conscious mental state ( $N_{ST}$ ). First, it is worth emphasising that in a proposition like (n), ‘Mary’ refers to a *subject of experience*, namely the subject of the token conscious mental state of toothache that is being talked about. While one might want to claim that subjects of experience are identical with *creatures* – or indeed, as Nagel would have it, *organisms* – this identity claim is presumably not true *a priori*. Thus, one might hold the view that subjects of experience are identical to proper parts of creatures, such as patterns of neural activity or information flow, or even that they are non-physical entities. This is a matter of substantive assumptions about the metaphysics of subjects of experience. However, we need not make such substantive assumptions to elucidate the meaning of consciousness. Thus, I will subsequently refer to *subjects* rather than *creatures* (or *organisms*) when discussing the Nagelian dictum and its variants. Accordingly, one can reformulate ( $N_{ST}$ ) as follows:

( $N_{ST1}$ ) A mental state *M* of a subject *S* is a *conscious* mental state of *S* at time *t* if and only if there is something it is like for *S* to be in *M* at *t*.

<sup>11</sup>See Stoljar (see 2016, p. 1176).

Second, when  $(N_{ST1})$  says, of a subject  $S$  and a mental state  $M$ , that there is something it is like *for*  $S$  to be in  $M$ , this ‘for’ should be understood as ‘for<sup>EXP</sup>’ – as when Nagel himself writes: “there is something that it is like to *be* that organism – something it is like *for* the organism” (Nagel 1974, p. 436). Given that the subject of the infinitive verb ‘to be (in  $M$ )’ is implicit, co-reference is required. Thus, the maximally explicit equivalent form is:

$(N_{ST2})$  A mental state  $M$  of a subject  $S$  is a *conscious* mental state of  $S$  at time  $t$  if and only if there is something it is like for<sup>EXP</sup>  $S$ , for<sup>SUBJ</sup>  $S$  to be in  $M$  at  $t$ .

And, in accordance with the affective account of the propositions that are stereotypically expressed by sentences containing the SIL phrase or the WIL phrase, we can unpack the right-hand side of  $(N_{ST2})$  as follows:

$(N_{ST3})$  A mental state  $M$  of a subject  $S$  is a *conscious* mental state of  $S$  at time  $t$  if and only if  $S$ ’s being in  $M$  at  $t$  affects  $S$  in some experiential way at  $t$ .

Third, Stoljar’s affective account allows that there may be something that it is like for<sup>EXP</sup> Mary for<sup>SUBJ</sup> Mary to have a hammer fall on her (unanaesthetised) foot, provided only that Mary is affected *in some experiential way* by having a hammer fall on her foot. (Indeed, Mary might be affected emotionally, and so experientially, even by a hammer falling on her anaesthetised foot.) However, having a hammer fall on her foot is not itself a conscious mental state (nor even a non-conscious mental state) of which Mary is the psychological subject.

Furthermore, the affective account also allows that there may be something that it is like for<sup>EXP</sup> Mary for<sup>SUBJ</sup> Mary to be in a *non-conscious* mental state, provided only that Mary’s being in that non-conscious mental state affects her in some experiential way. For example, Mary’s non-conscious desire for ice-cream might result in her feeling hunger, and thus affect Mary in an experiential way. In such a case,  $(N_{ST3})$  entails that Mary’s desire for ice-cream is in fact conscious; but this is not the notion of consciousness that the Nagelian dictum aims to capture. A conscious mental state must be *constitutively* such that it affects its subject in an experiential way.

Thus, what it is needed instead of  $(N_{ST3})$  is:

(N<sub>ST4</sub>) A mental state M of a subject S is a *conscious* mental state of S at time *t* if and only if M is *constitutively* such that S's being in M at *t* affects S in some experiential way at *t*.

As Stoljar puts it:

Nagel's definition requires... that if M is a *conscious* state then it is not simply that its subject must be affected in a certain way by being in M but moreover that this is in some way *essential or constitutive* of M.

Stoljar (2016, (p. 1190, emphasis added and 'X' replaced by 'M'))

Fourth, as I have noted, Stoljar uses “There is some way that S *feels* as a result of [or: in virtue of] S's being in M” as an alternative to “S's being in M affects S in some experiential way”. As Stoljar himself acknowledges, it might be objected that the concept of feeling does not encompass all the possible ways that a subject might be affected experientially. Stoljar suggests that one possible response to this objection would be “to insist on a concept of *feeling* broader than that at issue in the objection” (Stoljar 2016, p. 1181). Alternatively, we could use the broader verb ‘to experience’, or even coin a new verb ‘to *phenom*’ stipulated to express a broader concept, and deploy it in a version of (N<sub>ST</sub>) similar to Stoljar's (N5) (p. 1190):

(N<sub>ST5</sub>) A mental state M of a subject S is a *conscious* mental state of S at time *t* if and only if M is *constitutively* such that there is some way that S *phenoms* at *t* in virtue of being in M at *t*.

However, no new coining is needed. Instead, I shall supplement the Nagelian elucidation of consciousness as a property of mental states with a bridging principle (SIL) elucidating the meaning of the SIL phrase in terms of a mental state M being *constitutively* such a subject's being in M contributes to the subject's *overall phenomenology*:

(SIL) There is *something it is like* for a subject S to be in a mental state M at time *t* if and only if S's being in M at *t* *constitutively* contributes to S's overall phenomenology at *t*.

This demanding notion of there being something it is like for a subject to be in a mental state is arguably present in Nagel:

If mental processes are indeed physical processes, then there is something it is like, *intrinsically*, to undergo certain physical processes. What it is for such a thing to be the case remains a mystery.

[fn. 11] The relation would therefore not be a contingent one, like that of a cause and its distinct effect. It would be *necessarily true* that a certain physical state felt a certain way.

Nagel (1974, pp. 445-6, my emphasis)

Indeed, in the bridging principle (SIL), the relevant notion of a mental state's *constitutive* contribution to a subject's phenomenology is intended to contrast with a merely *causal* and *contingent* contribution to the subject's phenomenology. At a first pass, a mental state *constitutively* contributes to a subject S's overall phenomenology at *t* if and only if the following counterfactual obtains: if S were not in M at *t*, what it is like to be S at *t* would be different *ipso facto*.

Importantly, the inclusion of '*ipso facto*' in this counterfactual is intended to guarantee that it is *not* satisfied if M's contribution to S's overall phenomenology is merely causal. Suppose that you are (consciously) tasting orange juice. If you were experiencing the acid taste of lime juice rather than the sweet taste of orange juice, this would not merely *cause* a difference in your overall phenomenology; rather, this would precisely *be* the difference in your overall phenomenology. Experiencing the acid taste of lime juice would be an aspect of your overall phenomenology, rather than experiencing the sweet taste of orange juice.

Taken together, (N<sub>ST1</sub>) and (SIL) straightforwardly entail:

(N<sub>ST6</sub>) A mental state M of a subject S is a conscious mental state of S at time *t* if and only if S's being in M at *t* *constitutively* contributes to S's overall phenomenology at *t*.

(N<sub>ST6</sub>) satisfies the requirements I have set for the interpretation of the Nagelian dictum: it does not rely on special technical terms, and it does not express a comparative statement. As such, it provides the desired elucidatory account of (phenomenal) consciousness as a property of mental states.

Critics of the Nagelian dictum might object that none of the various versions of (N<sub>ST</sub>) I have distinguished are non-circular or “true definitions” (Chalmers 1996, p. 3) of consciousness as a property of mental states. Indeed, the relevant meanings of terms such as ‘experiential’, ‘to feel’ and ‘overall phenomenology’ cannot be understood independently of the meaning of ‘conscious’. Thus, the Nagelian account does not define consciousness in more fundamental terms, and does not provide an explanatory account of consciousness. This why I have referred to it as an ‘elucidation’ rather than a ‘definition’ strictly speaking.

What the Nagelian account does provide, however, is a non-technical way to grasp what the obvious folk-psychological concept that ‘conscious’ refers to, when it is predicated of mental states. Ultimately, there is no way to capture this concept other than circling around it with paraphrases; but this is not problematic given the assumption that subject endowed with suitable rational capacities possess the relevant concept, and can successfully latch onto it when reading a claim such as (N<sub>ST6</sub>).

Furthermore, (N<sub>ST6</sub>) elucidates the nature of the relationship between conscious mental states and psychological subjects. Indeed, it makes it clear that: (a) a mental state’s being a *conscious* mental state involves that mental state being related in an appropriate way to a subject; (b) the appropriate way in which a *conscious* mental state is related to its subject is an *experiential way* of affecting its subject; (c) finally, a *conscious* mental state is *constitutively* such that it affects its subject in an experiential way, that is, a *conscious* mental state *constitutively* contributes to its subject’s overall phenomenology.

## 1.5 The phenomenal character of consciousness

My discussion of the Nagelian dictum has mostly focused on the SIL phrase as it is used in (SIL) and the various versions of (N<sub>ST</sub>). In these claims, the SIL phrase is used to express the idea that a mental state is a *conscious* mental state of a subject, rather than a *non-conscious* mental state of a subject (or, equivalently, that a mental state is constitutively such that it makes *some* contribution to its subject’s overall phenomenology, rather than *no* contribution at all). This is what the Nagelian dictum sought to elucidate.

While the meaning of the WIL phrase is intimately related to that of the SIL phrase, there is an important difference between the two phrases. The SIL phrase is used to express

an existentially quantified claim – there is *something*, rather than *nothing*, that it is like for a subject S to be in a conscious mental state M. By contrast, the WIL phrase is used to refer to the ‘witness’ (in the logical sense) of such an existential claim – provided that there is *something*, rather than *nothing*, that it is like for S to be in M, the WIL phrase refers to *what* S’s being in M is like. In other words, the WIL phrase does not merely capture the fact that M makes *some* contribution or other to S’s phenomenology, but rather *what* M’s contribution to S’s overall phenomenology is.

By analogy with the bridging principle (SIL), we can thus formulate a bridging principle (WIL) to elucidate the meaning of the WIL phrase as follows:

(WIL) *What it is like* for a subject S to be in a mental state M at time *t* is what S’s being in M at *t* constitutively contributes to S’s overall phenomenology at *t*.

We are now well-equipped to distinguish between two properties captured respectively by the SIL and WIL phrases. The SIL phrase captures the *second-order* property of making *any* contribution at all to the subject’s overall phenomenology, rather than *none* at all. Let us call this second-order property ‘*phenomenality*’. When a mental state is constitutively such that there is something it is like for its subject to be in it, that mental state instantiates the second-order property of phenomenality. In other words, phenomenality refers to the feature that is common across *all* conscious mental states (and that all non-conscious mental states lack).

In turn, the WIL phrase captures the first-order property of making *a specific* contribution to the subject’s overall phenomenology, rather than *some other specific* contribution. This first-order property is commonly called the *phenomenal character* of a conscious mental state. The phenomenal character of a conscious mental state is *what* that mental state constitutively contributes to the subject’s overall phenomenology. It can be conceived as the sum of the mental state’s *phenomenal properties* (such as the sharpness and the vividness of a painful sensation).

Note that the conceptual distinction between phenomenality and phenomenal character is apparent from attempts to define consciousness by example: the experience of tasting a sweet orange and the experience of seeing a yellow lemon are both *conscious* (rather than

*non-conscious*) mental states, in so far as there *something* (rather than *nothing*) that it is like to taste an orange and to see a lemon. Thus, they both instantiate *phenomenality*. However, *what* it is like to taste an orange differs from *what* it is like to see a lemon, in so far as these experiences make distinct contributions to the subject's overall phenomenology. Thus, the experience of tasting a sweet orange and the experience of seeing a yellow lemon differ with respect to their phenomenal character, but not with respect to their phenomenality.

As we shall see in chapter 3, it is important not to get confused about the distinction between phenomenality and phenomenal character. Any mental state that instantiates the first-order property of having a certain phenomenal character *ipso facto* instantiates the second-order property of phenomenality, as the latter merely involves existential quantification over the former. Consequently, it would be double-counting to treat the second-order property of phenomenality as if it were a *further* first-order property to be specified as part of a mental state's phenomenal character. As a matter of logic, rather than substantive philosophical theory, a specification of *what* it is like to be in a mental state M *guarantees* that there is *something* it is like to be in M; conversely, the fact that there is *something* it is like to be in M *guarantees* that there is some answer to the question of *what* it is like to be in M. If there is *nothing* it is like to be in M, then the question of *what* it is like to be in M does not arise.

As a simple example of this distinction between first-order and second-order properties, consider the case of coloured objects. Some worldly objects are coloured – they have some colour or other, rather than none. We could use the term '*chromaticity*' for the second-order property of having *some* colour(s) or other, rather than none; and we could use the term '*chromatic character*' for the first-order property of having *some specific* colour(s), rather than any other(s). Accordingly, all coloured objects instantiate chromaticity, although they may differ with respect to their chromatic character. However, we cannot treat chromaticity as if it were a *further* first-order property of coloured objects to be specified as part of the object's chromatic character.

The relationship between phenomenality and phenomenal character, just as the relation between chromaticity and chromatic character, can also be glossed as a *determinable-determinate* relation. Having chromaticity (i.e., being coloured at all) is a *determinable*

property of which having a certain chromatic character (e.g., being red) is a *determinate*. Likewise, having phenomenality (i.e., making any constitutive contribution to the subject phenomenology) is a *determinable* property of which having a specific phenomenal character (e.g., being painful) is a *determinate*.<sup>12</sup>

With this elucidatory account of consciousness in mind, I will now consider which foundational claims about consciousness plausibly follow from this account.

## 1.6 Foundational claims about consciousness

### 1.6.1 Claims about the (self-)ascription of conscious mental states

Recall that we have previously captured the spirit of the Nagelian dictum, under the affective interpretation, in the following claim:

(N<sub>ST6</sub>) A mental state *M* of a subject *S* is a *conscious* mental state of *S* at time *t* if and only if *S*'s being in *M* at *t* *constitutively* contributes to *S*'s overall phenomenology at *t*.

A few claims plausibly follow from this elucidation of consciousness. First, we can link (N<sub>ST6</sub>) to consciousness as a property of subjects (or creatures). To refer to such a property, Nagel uses the phrases ‘something it is like for an organism *to be that organism*’ and ‘what it is like for an organism *to be that organism*’. These formulations are potentially confusing, for they seem to invite the idea that there is something it is like to *be* Mary, which could be different from what it is like to *be* Paul, even if Mary and Paul are in the same type of total phenomenal state.<sup>13</sup> To prevent this misinterpretation, I will use the locutions ‘something it is like for *S*’ and ‘what it is like for *S*’ without the addition of the embedded clause ‘to be *S*’ (this is slightly awkward in English, but less likely to invite misinterpretation).

It follows from what precedes that an account of what it is like for a subject *S* at time *t* is equivalent to an account of *S*'s overall phenomenology (or total phenomenal state) at

<sup>12</sup>See also Loar (1990, p. 95) and Kriegel (2015, p. 10) on the determinable-determinate relations between phenomenal properties.

<sup>13</sup>Nagel himself guards against this interpretation when he says “I am not adverting here to the alleged privacy of experience to its possessor” (Nagel 1974, p. 441).

*t*. Furthermore, it plausibly follows from (N<sub>ST6</sub>) that any account of what it is like for S at *t* must refer to the conscious mental state(s) that S is in at *t*:

(F1) If M is a conscious mental state of subject S at *t*, then an account of what it is like for S at *t* (i.e., an account of S's overall phenomenology at *t*) must include that S is in M.<sup>14</sup>

An *account* of what it is like for you – that is, an account of your overall phenomenology – while you are (consciously) tasting orange juice must include *that you are experiencing the taste of orange juice*. But it is important to distinguish between a subject's being in a conscious mental state and giving an account of what it is or was like for the subject to be in that conscious mental state. Your being in conscious mental state M at time *t* does not depend on you (or anyone else) giving an account at *t* (or at any other time) of your being in M.

Importantly, the following claim does *not* follow from (N<sub>ST6</sub>) or (F1):

(S1) If M is a conscious mental state of subject S, then S is in a conscious mental state M\* with the content <S [SELF] is in M>.<sup>15</sup>

That a conscious mental state constitutively contributes to a subject's phenomenology does not entail that the subject in question should be in a conscious mental state – whether a distinct, higher-order conscious mental state or the very same first-order conscious mental state – whose content is a self-ascription of the first-order conscious mental state. Thus, (S1) is a substantive claim that needs to be motivated independently of (N<sub>ST6</sub>) and (F1), whereas (F1) itself flows quite naturally from the Nagelian account of a conscious mental state.

Toward the end of his paper on the Nagelian account, Stoljar (2016) considers whether the Nagelian definition of a conscious mental state entails a claim that is similar to (S1); namely “an individual is in a conscious state only if the individual represents or is aware of (in some sense) their being in that state” (p. 1193). Stoljar finds this claim in several

<sup>14</sup>For clarity, I will flag foundational claims that plausibly follow from an elucidatory account of consciousness with an ‘F’ for ‘foundational’.

<sup>15</sup>For clarity, I will flag the substantive claims that do *not* plausibly follow from an elucidatory account of consciousness with an ‘S’ for ‘substantive’. I use angle brackets to denote the content of mental states.

philosophers (e.g., Kriegel 2009; Levine 2007; Weisberg 2011) and he formulates a general argument that seems to lie behind their claim. Stoljar calls this the ‘emphatic argument’ because it proceeds by placing emphasis on the ‘for S<sup>EXP</sup>’ in the Nagelian definition:

- (P1) M is a phenomenally conscious state of S only if there is something it is like *for* S to be in M.
- (P2) There is something that it is like *for* S to be in M only if S stands in some representational or awareness relation to M.
- (C) M is a phenomenally conscious state of S only if S stands in some representational or awareness relation to M.

(Stoljar 2016, pp. 1193-4; with ‘X’ replaced by ‘M’)

The first premise (P1) is indeed a consequence of the Nagelian definition, but that definition (with its ‘something it is like’ phrase) has been elucidated in terms of Stoljar’s affective account. The formulation that Stoljar discusses (p. 1194) is his (N5) but the two points that he makes about the emphatic argument are equally clear if we consider, for example, (N<sub>ST4</sub>) or (N<sub>ST6</sub>). The first point is that there are versions of the first premise that are plausible consequences of the Nagelian definition – for example (P1\*) based on (N<sub>ST4</sub>), or (P1\*\*) based on (N<sub>ST6</sub>) – but the phrase ‘for S’ does not appear in these formulations; “so nothing at all may be wrung from this phrase in particular” (p. 1194):

- (P1\*) M is a conscious mental state of S only if M is constitutively such that S’s being in M affects S in some experiential way.
- (P1\*\*) M is a conscious mental state of S only if S’s being in M constitutively contributes to S’s overall phenomenology.

The second point is that the second premise that is then required will be: “unpersuasive to someone who does not already agree with [the claim (C) that is being argued for]” (ibid.). Thus, if we modify the second premise as we have modified the first premise – for example (P2\*) based on (N<sub>ST4</sub>), or (P2\*\*) based on (N<sub>ST6</sub>) –, we can see that this premise requires independent motivation:

- (P2\*) M is constitutively such that S’s being in M affects S in some experiential way only if S stands in some representational or awareness relation to M.

(P2\*\*) S's being in M constitutively contributes to S's overall phenomenology only if S stands in some representational or awareness relation to M.

Thus, following Stoljar (2016), we reach the conclusion that the emphatic argument does not provide a persuasive route from the Nagelian definition of a conscious mental state to a claim similar to (S1). This is not to say that claim (S1) is false, but only that it requires its own motivation. I shall come back to the emphatic argument in chapter 3.

Are there foundational claims about consciousness in the vicinity of (S1) that, unlike (S1) itself, plausibly follow from an elucidation of consciousness? There is an air of resemblance between (S1) and so-called constitutive accounts of self-knowledge. According to such accounts, a subject's having a first-order mental state M depends on her believing that she is in M, in so far as this belief is constitutive of the first-order mental state. Thus, on such accounts, "the existence of a second-level belief about a first-level psychological state is itself what makes it true that the first-level state exists." (Heal 2002, p. 4).<sup>16</sup> However, constitutive accounts of self-knowledge do not entail that in order to be in a first-order mental state M, one has to *consciously* believe that one is in M. As Boyle (2011) puts it, speaking of cases in which the first-order mental state is a belief, "it may be true that a person's believing P involves his knowing himself to believe P [as the constitutive account claims], and yet that a person can believe P *without being conscious of it*" (p. 230, my emphasis). Boyle's remark generalises to any first-order mental state (beyond first-order beliefs), including any *conscious* first-order mental state. Consequently, constitutive accounts of self-knowledge do not entail a claim like (S1) without further assumptions that need not be part of such an account. Furthermore, the assumptions on which constitutive accounts of self-knowledge do typically rely are themselves substantive claims that, like (S1), do not plausibly follow from the Nagelian account of consciousness.<sup>17</sup>

A less controversial claim about our knowledge of our conscious mental states is that subjects are, generally, *in a position to know* that they are in a conscious mental state just

<sup>16</sup>See also Shoemaker (1994b, p. 272), Zimmerman (2006, p. 338), and Boyle (2011, p. 235).

<sup>17</sup>See Parrott (2017) for a critical discussion of the controversial assumptions of constitutive accounts of self-knowledge.

by virtue of being in it. At a first pass, this claim seems to be threatened by Williamson's anti-luminosity argument (Williamson 2000). In Williamson's terms, a given condition is *luminous* if and only if it is such that *whenever it obtains*, one is in a position to know that it obtains. Williamson's anti-luminosity argument purports to show that there is "virtually no mental state" such that the condition that one is in that mental state is a luminous condition (Williamson 2000, p. 14). While the argument focuses on establishing that the condition *that one feels cold* is not luminous, it is supposed to generalise to almost every mental state, including conscious mental states. Williamson's initial example (2000) involves a subject S feeling extremely cold at dawn, and very slowly warming up until she feels hot at noon. At some time  $t_1$  during this gradual transition, S *barely* feels cold, and still truly believes that she feels cold; but at time  $t_2$ , one millisecond later, she feels ever so slightly warmer, and thus no longer actually feels cold.

According to Williamson, in order for S's belief that she feels cold at  $t_1$  to be sufficiently safe from error to constitute knowledge, it must be the case that this belief is not false in any similar situation that S cannot discriminate from what is the case at  $t_1$ . However, S's situation at  $t_2$  – a mere one millisecond later – is precisely such a situation, for S cannot possibly discriminate between how she feels at  $t_1$  and how she feels at  $t_2$ ; yet, at  $t_2$ , S's belief that she feels cold is false. Therefore, Williamson concludes that S is not really in a position to know that she feels cold at  $t_1$ , because at that time her true belief that she feels cold is not sufficiently sheltered from the possibility of error to constitute knowledge. It follows that the condition *that one feels cold* is not luminous, because there are cases in which one feels cold without being in a position to know that one does. From there, Williamson argues that this argument schema can generalise to almost any condition, such as being in pain, feeling tired or having a headache.

Let us assume that the anti-luminosity argument successfully shows that there are a number of conscious mental states M such that the condition *that one is in M* is not a luminous condition. This would entail that subjects are not *always* in a position to know that they are in a given conscious mental state just by virtue of being in it. However, the anti-luminosity argument does not threaten two weaker claims about subjects' epistemic and doxastic attitudes towards their own conscious mental states. The first of these claims

focuses on justified belief rather than knowledge, and thus avoids Williamson's strong constraint on safety from error:

(F2) If M is a conscious mental state of subject S, then S is in a position to justifiably believe that S [SELF] is in M.

This claim borrows from Berker (2008), who coins the notion of a *lustrous* condition to characterize a condition such that one is in a position to *justifiably believe* that it obtains whenever it does. Berker convincingly argues that an argument schema analogous to the anti-luminosity argument would not be successful at establishing that the condition *that one feels cold* is not lustrous, because justified belief does not require safety from error. For the subject of Williamson's thought experiment to justifiably believe that she feels cold at  $t_1$  (when she *barely* feels cold), it need not be the case that she actually feels cold and her belief remains true a millisecond later at  $t_2$ . Indeed, she would be perfectly justified in believing that she feels cold at  $t_1$  even though (a) this belief would be false at  $t_2$  and (b) she would not be able to discriminate between her situation at  $t_1$  and her situation at  $t_2$ . The same remark applies to any other mental state: there is presumably no mental state M such that an argument analogous to the anti-luminosity argument could show that the condition *that one is in M* is not lustrous. If this is correct, then (F2) is plausible even if the anti-luminosity argument is successful. Suppose that you are (consciously) tasting orange juice. It follows from (F2) that you are in a position to justifiably believe that you are experiencing the taste of orange juice. Specifically, you are in a position to rationally form the belief that you are having an experience which has the phenomenal character it has – an experience *as of* tasting orange juice.

If (F2) plausibly follows from an elucidatory account of consciousness, then so does the claim that being in a conscious mental state gives one a reason to judge rationally that one is that state:

(F3) If M is a conscious mental state of subject S, then S's being in M gives S a reason to judge rationally that S [SELF] is in M.

Another plausible claim that one can make even if the anti-luminosity argument succeeds in showing that some conditions are not luminous is that, *in most cases* in which one is in a given conscious mental state, one is in a position to know that one is in that state. Indeed, Williamson's example of the gradual transition from feeling cold to feeling warm involves a rather unusual situation in which one barely feels cold at one moment and no longer feels cold at the next. There are less marginal cases, for example, cases in which one feels *very* cold. In such cases, one's belief that one feels cold would not be false in a nearly identical case, and would thus be plausibly safe from error in Williamson's sense (DeRose 2002). This point generalises to most of our conscious mental life, as our justified beliefs about the conscious mental states that we are in are quite often sheltered from the kind of error illustrated by Williamson's example. In most cases, subjects are not only in a position to make rational judgements and adopt rational beliefs about their own conscious mental states, but they are also presumably in a position to *know* about their conscious mental states. Williamson's anti-luminosity argument shows only that there are some cases in which we are not in a position to know about our (conscious) mental states. Thus, the following claim appears to be very plausible:

(F4) *In most cases*, if M is a conscious mental state of a subject S, then S is in a position to know that S [SELF] is in M.

If you are (consciously) tasting orange juice, having that experience puts you in a position to know, at least in most cases, that you are experiencing the taste of orange juice (i.e., that you are having an experience which has the character of tasting orange juice).<sup>18</sup> Indeed, you are in a position to have a justified belief that you are experiencing the taste of orange juice (by F2), and in a typical case such belief is reasonably safe from error (it would not be false in a nearly identical case that you could not discriminate from *c*, such as a case in which you experience an ever so slightly sweeter taste).

<sup>18</sup>Of course, this does not entail that you are in a position to know that the liquid you are tasting is, in fact, orange juice (or indeed that there is any liquid that you are in fact tasting, for you could be hallucinating the taste of orange juice).

Importantly, there is no persuasive route from ( $N_{ST6}$ ) or from any or all of the foundational claims (F1), (F2), (F3), and (F4) to (S1). Indeed, what puts a subject S in a position rationally to believe and judge, and usually to know, that she (*herself*) is in a conscious mental state M need not be that she is already in a conscious mental state  $M^*$  with the content  $\langle S$  [SELF] is in M $\rangle$ . Rather, what enables S to form such doxastic attitudes and to acquire such knowledge is just that M is a conscious mental state of which S is the subject – or, to put it another way, that M constitutively contributes to S’s overall phenomenology. Thus, endorsement of (S1) must depend on further substantial commitments.

### 1.6.2 Claims about the subject of experience

Two other important foundational claims about consciousness plausibly follow from the Nagelian account. These claims relate to the relationship between mental states and subjects of experience. The first is that every conscious mental state has a subject of experience, or as Frege puts it:

It seems absurd to us that a pain, a mood, a wish should rove about the world without a bearer, independently. There is no experience without a subject of experience. The inner world presupposes the person whose inner world it is.<sup>19</sup>

Frege (1921/1956, p. 299, translation modified)

More precisely:

(F5) If M is a conscious mental state, then there is a subject S such that S is in M (and, consequently, S’s being in M constitutively contributes to S’s overall phenomenology).

Any episode of experiencing is undergone – enjoyed or endured – by a subject of experience. A mental state is a conscious mental state if and only if there is something it is like to be in it; and it is only *for someone* – for a subject of experience – that there can be something it is like to be in a mental state. Only a subject of experience has an *overall phenomenology* to which being in a mental state contributes; only a subject of experience can be *affected in an experiential way* by being in a mental state.

<sup>19</sup>I have modified A.M. and Marcelle Quinton’s original translation, “An experience is impossible without an experient”. See also Shoemaker (1986, p. 10).

A distinct foundational claim about subjects of experience is that no conscious mental state can have more than *one* subject of experience. In other words, for any given (token) conscious mental state *M*, there is no more than *one* subject of experience who is in *M* and for whom there is something it is like to be in *M*:

(F6) There is no (token) mental state *M* such that *M* constitutively contributes to the overall phenomenology of more than one subject.

A given experience cannot be undergone by several subjects. If I am in pain, only I can feel *that specific episode of pain*. Furthermore, this is a matter of metaphysical necessity, and thus it should not be possible to find counter-examples to (F6).

Upon observing someone being touched on a specific area of the body (e.g. on the index finger), individuals with vision-touch synaesthesia (also called mirror-touch synaesthesia) routinely feel a tactile sensation on the same area (e.g. on the index finger) of their own body (Ward and Banissy 2015). There is a metaphorical sense in which vision-touch synaesthetes may 'feel' what another subject feels upon being touched. However, even such synaesthetes do not literally have the same token experience as the observed individual.<sup>20</sup> At best, they have the same *type* of experience, because the observation of a tactile interaction triggers for the synaesthete a tactile sensation of same type as the tactile sensation that the seen touch causes (or would normally cause) in the observed individual.

The impossibility of sharing token conscious mental states can be explained by the fact that mental states, and specifically conscious mental states, are individuated not only by their type but also by their psychological subject. In other words, a mental state's subject features among the mental state's conditions of individuation as a matter of metaphysical necessity. Consequently, token experiences can neither be *free-floating* (F5) nor *shared* (F6): they make a constitutive contribution to *one and only one* subject's overall phenomenology.

<sup>20</sup>In fact, the seen subject's actual experience is irrelevant to the experience of vision-touch synaesthetes. For example, if a synaesthete sees someone being touched on the arm while the arm is in fact anaesthetised, that will not prevent the synaesthete from experiencing a tactile sensation on her own arm. There is even evidence that vision-touch synaesthesia can be triggered by seeing a rubber hand being touched (Aimola Davies and R. C. White 2013).

I take it that the claims (F1), (F2), (F3), (F4), (F5) and (F6) are relatively uncontroversial given the Nagelian elucidation of a conscious mental state ( $N_{ST6}$ ). In the next chapter, I will distinguish between two broad concepts of 'self-consciousness', which in turn yield two versions of the claim that self-consciousness is constitutive of consciousness.

## 2

# Two Concepts of Self-Consciousness

The term 'Self-Consciousness' opens up a very wide discussion... The name Consciousness standing single [is] intelligible and free from ambiguity. The addition of the prefix 'Self' entirely changes the situation.

Bain (1894, p. 358)

In what precedes, I offered an elucidatory account of consciousness, and I outlined several foundational claims that plausibly follow from such an account. I shall now turn to the notion of self-consciousness, and to the claim that self-consciousness is constitutive of consciousness.

Self-consciousness is a form of consciousness; in other words, an instance of self-consciousness is, *ipso facto*, an instance of consciousness. More specifically, self-consciousness is a form of consciousness construed *as a property of subjects*. Indeed, one does not typically say of a *mental state* that it is 'self-conscious'; but one does typically say of a *subject* that he or she is 'self-conscious'. Thus, a self-conscious subject is, *ipso facto*, a conscious subject – that is, a subject who is in a mental state that makes a constitutive contribution to his or her overall phenomenology. However, it does not follow that a conscious subject is, *ipso facto*, a self-conscious subject; more needs to be said to understand in what sense self-consciousness might be constitutive of consciousness.

In §2.1, I distinguish two broad concepts of self-consciousness, namely (a) consciousness of *consciousness itself*, and (b) consciousness of *oneself*. While these concepts are occasionally conflated, they are not *prima facie* equivalent. In §2.2, I formulate two broad variants of the claim that consciousness constitutively involves self-consciousness, each of which makes use of one of the two concepts of self-consciousness previously distinguished. Finally, in §2.3, I review the surprisingly long history of each of the two resulting claims,

and emphasise some questions raised by this historical survey regarding contemporary versions of the claims.

## 2.1 Defining self-consciousness

Defining the notion of self-consciousness is a matter of specifying how the prefix *self-* qualifies the meaning of ‘consciousness’. However, this prefix can have several grammatical functions, and this is an important source of ambiguity in the meaning of ‘self-consciousness’.

### 2.1.1 Self-consciousness as consciousness *of consciousness itself*

A first notion of self-consciousness results from the somewhat unorthodox use of the prefix *self-* to indicate that the root of the lexical compound (‘consciousness’) is directed at itself or takes itself as its own object.<sup>1</sup> Accordingly, in a philosophical context, the notion of self-consciousness can refer to consciousness *of consciousness itself*. As we shall see in section 2.3, this use of the prefix *self-* in ‘self-consciousness’ was explicitly introduced by Sartre (1943/1948) in French (*conscience de soi*’ being defined as *conscience de conscience*’), although one can find similar formulations in earlier works.

The notion of ‘consciousness *of consciousness itself*’ stands in need of further elucidation. At a first pass, it refers to a subject’s consciousness of her own consciousness:

- (1) A subject S is *self-conscious* at *t* if and only if S is conscious at *t* and S is conscious of *that consciousness itself* at *t*.

Of course, (1) does not clarify what it means for a conscious subject to be conscious *of her own consciousness*. Given that a subject is conscious at a time if and only if she is in a conscious mental state at that time, we can plausibly interpret (1) as follows:

- (2) A subject S is *self-conscious* at *t* if and only if S is in a conscious mental state M at *t* and S is conscious *of M* at *t*.

<sup>1</sup>This use of the prefix *self-* is equivalent to the more prevalent (although recent) use of the prefix *meta-* (e.g. *meta-data* is data about data).

Note that a more complex and less straightforward interpretation of (1) is available, on which a subject being conscious of her own consciousness is equivalent to her being conscious of *her being conscious* or of *her being in a conscious mental state*:

(3) A subject S is *self-conscious* at *t* if and only if S is conscious at *t* and S is conscious of S's *being conscious* at *t*.

(4) A subject S is *self-conscious* at *t* if and only if S is in a conscious mental state M at *t* and S is conscious of S's *being in M* at *t*.

It is natural to understand the phrase '*x* is conscious of *being F*' as shorthand for '*x* is conscious *that x is F*'. On such a reading, (3) and (4) state that a subject being self-conscious is a matter of her being conscious *that she* [SELF] *is conscious*, or *that she* [SELF] *is in a conscious mental state*. However, what this propositional use of 'being conscious' would mean is unclear, as consciousness is not typically taken to be a propositional attitude.<sup>2</sup> Furthermore, defining self-consciousness as the subject's consciousness *that she* [SELF] *is conscious* or *that she* [SELF] *is in a conscious mental state* seems quite demanding, as such a notion of self-consciousness would presumably require the use of a concept of self. On the face of it, no such use is required for a subject to be self-conscious in the sense defined by (2).

As we shall see, most authors who argue that self-consciousness is constitutive of consciousness do not endorse the claim that self-consciousness involves a propositional attitude, or requires the use a concept of self (for this would entail, rather implausibly, that being conscious at all requires bearing a propositional attitude or using a concept of self). Thus, most authors who argue that consciousness constitutively involves consciousness of *consciousness itself* understand the latter as the subject's consciousness of *her occurrent conscious mental state* – as defined by (2).

<sup>2</sup>It is possible, although perhaps not particularly idiomatic, to use consciousness as a propositional attitude in some colloquial contexts – for example, one could say "Mary is conscious *that it is raining*". However, this sentence would normally express an epistemic fact, namely the fact that Mary *knows* that it is raining. This has little to do with the notion of consciousness elucidated in chapter 1.

### 2.1.2 Self-consciousness as consciousness of oneself

A more common category of lexical compounds formed with the prefix *-self* results from the use of the prefix to indicate that the *subject* of the act or attitude denoted by the root of the compound is also the *object* of that act or attitude. For example, *self-love* denotes the love that a subject has *for that same subject* (i.e., *for herself*); *self-deception* denotes the fact that a subject deceives *that same subject* (i.e., *herself*); and an object is *self-adhesive* if it adheres *to itself*.

Understood as a member of this category, ‘self-consciousness’ refers to the subject’s being conscious of *that same subject* (i.e., *of herself*):

(5) A subject S is *self-conscious* at *t* if and only if S is conscious of S [SELF] at *t*.

However, there is a lingering ambiguity in (5), for one can be conscious of (what is in fact) oneself either *accidentally* or *non-accidentally*. For example, one might have a conscious thought about the person who left the fridge door open, or see a man standing in the hallway through a mirror, without being conscious of that person *as oneself* – when that person in fact happens to be oneself. In such scenarios, by consciously thinking of the person who left the fridge door open, or by consciously seeing the man standing in the hallway, one is *de facto* conscious of oneself in a merely accidental way.<sup>3</sup> In the philosophical literature, the notion of *consciousness of oneself* typically refers to cases in which one is *essentially* or *non-accidentally* conscious of oneself. Being self-conscious, on this account, is not simply being conscious of an *x* such that *x* happens (as a matter of fact) to be oneself; it is, rather, being conscious of an *x as oneself*.

To capture this nuance, we can understand (5) as follows:

(6) A subject S is *self-conscious* at *t* if and only if S is conscious of S *as oneself* at *t*.

This definition of self-consciousness raises a further question: does being conscious of oneself *as oneself* require the possession and use of a concept of self? While *thinking of*

<sup>3</sup>I will come back to the distinction between being conscious of oneself *accidentally* and *non-accidentally* in chapter 5, which focuses on conscious thoughts about oneself.

oneself *as oneself* does seem to require a concept of self, the issue is less straightforward in other cases. Suppose, for example, that one can be conscious of oneself *as oneself* in virtue of being conscious of a body part *as a part of oneself*. As we shall see in chapter 6, being conscious of oneself as oneself in such a way might not involve a concept of self, but might merely involve instead a nonconceptual form of self-representation. If it is possible to represent oneself *as oneself* nonconceptually, that is without the use of a concept of self, then we should not conclude from (6) that self-consciousness – defined as consciousness of oneself *as oneself* – requires a concept of self.

### 2.1.3 A deflationary notion of self-consciousness?

I have so far considered two functions of the prefix *self-* as features in the lexical compound ‘self-consciousness’: (a) indicating a relation of consciousness to itself (i.e., self-consciousness construed as consciousness *of consciousness itself*), and (b) to denote the self as the object of consciousness (i.e., self-consciousness construed as consciousness *of oneself as oneself*). Perhaps one could construe the prefix *self-* in ‘self-consciousness’ as simply denoting the *subject* of consciousness – *whose* consciousness it is, rather than *what* or *whom* it is consciousness *of*. This would be the equivalent of construing the phrase ‘consciousness of self’ as a *subjective* genitive rather than an *objective* genitive. Take for example the phrase ‘love of God’; it can refer either to someone’s (anyone’s) love *for* God (objective genitive), or to *God’s own love* (subjective genitive). Similarly, the phrase ‘consciousness of Paul’ can refer either to someone’s consciousness *of* Paul (e.g., when someone sees Paul), or to *Paul’s own consciousness*. By the same token, the phrase ‘consciousness of self’ could refer to one’s consciousness *of oneself*, or – slightly awkwardly in English – to *one’s own consciousness*. The same reading could be applied, by extension, to ‘self-consciousness’. On this deflationary construal, ‘being self-conscious’ would mean ‘being *oneself* conscious’:

(7) A subject S is *self-conscious* at *t* if and only if S is (*herself*) conscious at *t*.

According to (7), ‘self-consciousness’ would be a mere synonym for consciousness, because a sentence such as “Mary is self-conscious” would be equivalent to “Mary is (*herself*) conscious”, or simply “Mary is conscious”.

Why should we care at all about this seemingly vacuous (and unidiomatic) understanding of self-consciousness? The answer is that some authors do seem to use the notion of self-consciousness as a mere synonym for consciousness. Consider, for example, the following definition of self-consciousness in a ‘weak’ or ‘minimal’ sense:

[A]ll subjective experience is self-conscious in the weak sense that there is something it is like for the subject to have that experience.<sup>4</sup>

Flanagan (1992, p. 194)

On this definition, being self-conscious (in the ‘weak’ or ‘minimal’ sense) simply is being conscious. Of course, if self-consciousness in one of the more specific senses outlined above (*consciousness of consciousness itself* or *consciousness of oneself*) turns out to be constitutive of consciousness, then it does follow that all instances of consciousness are, *ipso facto*, instances of self-consciousness. However, this would presumably be a substantive claim. By contrast, if self-consciousness merely denotes the consciousness of some subject of experience, the idea that all consciousness is self-consciousness might not be so substantive.

When philosophers argue for the claim that consciousness constitutively involves self-consciousness, they need to demonstrate that they have not slipped into the deflationary construal of self-consciousness expressed by (7). In chapters 3 and 4, I will assess whether arguments for the constitutive claim (on either of the two putatively substantial construals of self-consciousness) actually meet this standard. For now, we can leave the deflationary construal of self-consciousness aside, and focus on the two putatively substantial concepts of self-consciousness outlined above – namely (a) *consciousness of consciousness itself* and (b) *consciousness of oneself (as oneself)*.

#### 2.1.4 A note on terminology

Given the polysemy of the term ‘self-consciousness’, it is helpful to agree on some special terminology to easily discriminate the notion of *consciousness of consciousness itself* from

<sup>4</sup>Interestingly, Flanagan immediately goes on to defend a notion of self-consciousness that appears to go *beyond* the Nagelian account of consciousness: “This involves a sense that the experience is the subject’s experience, that it happens to her, occurs in her stream” (Flanagan 1992, p. 194). I will come back to this misleading semantic shift in chapters 3 and 4.

the notion of *consciousness of oneself (as oneself)*. Drawing upon Gurwitsch (1941), I will refer to the former notion as ‘non-egological self-consciousness’, and to the latter notion as ‘egological self-consciousness’. In the context of this distinction, ‘egological’ expresses the fact that the relevant kind of self-consciousness involves a form of *self*-representation – as representation of *oneself as oneself*. By contrast, ‘non-egological self-consciousness’ does not involve such self-representation.<sup>5</sup> One reason to adopt Gurwitsch’s terminology over alternatives is that it is the only option that comes close to being the standard way of drawing the distinction between the two notions of self-consciousness.<sup>6</sup>

## 2.2 Two constitutive claims: a first pass

We are now well-equipped to distinguish between two broad variants of the claim that consciousness constitutively involves self-consciousness. The generic version of the claim can be formulated as follows – where ‘CSC’ stands for ‘constitutiveness of self-consciousness’:

(CSC) Necessarily, for any subject S, if S is conscious at *t*, then S is *self-conscious* at *t*.

With the distinction between the non-egological and the egological notions of self-consciousness in mind, we can distinguish two variants (CSC<sub>NE</sub>) and (CSC<sub>E</sub>):

(CSC<sub>NE</sub>) Necessarily, for any subject S, if S is *conscious* at *t*, then S is *self-conscious* – in the *non-egological* sense – at *t*.

(CSC<sub>E</sub>) Necessarily, for any subject S, if S is *conscious* at *t*, then S is *self-conscious* – in the *egological* sense – at *t*.

As we have seen, the non-egological notion of self-consciousness (or consciousness of *consciousness itself*) is not typically intended to be a matter of bearing an attitude to the

<sup>5</sup>Note that Gurwitsch himself talks about the ‘non-egological conception of consciousness’ and the ‘egological conception of consciousness’ in his (1941). In my preferred terminology, the former refers to the claim that consciousness constitutively involves consciousness of consciousness itself (i.e., what I shall now call ‘non-egological self-consciousness’), while the latter refers to the claim that consciousness constitutively involves consciousness of oneself as oneself (i.e., what I shall now call ‘egological self-consciousness’).

<sup>6</sup>See, for example, Frank (2007, p. 154), Kriegel (2009, pp. 177-9), and Zahavi (2014, p. 48).

proposition that one is conscious of some item. Rather, it is more adequately captured by the subject's being conscious *of her conscious experience or mental state*. Consequently, we can reformulate (CSC<sub>NE</sub>) more specifically as follows:

(CSC<sub>NE</sub>) Necessarily, for any subject S and any mental state M of S, M is a *conscious* mental state of S at *t* if and only if S is conscious *of M* at *t*.

Note that in this claim, the notion of consciousness that is elucidated in terms of (non-egological) self-consciousness is consciousness as a property of mental states rather than subjects. However, given that a subject is *conscious* if and only if she is in a *conscious* mental state, we can easily adapt (CSC<sub>NE</sub>) to focus on consciousness as a property of subjects:

(CSC<sub>NE</sub>\*) Necessarily, for any subject S, S is *conscious* at *t* if and only if (a) S is in a conscious mental state M at *t* and (b) S is conscious *of M* at *t*.

In turn, we have defined the egological notion of self-consciousness as consciousness *of oneself*. We can thus reformulate (CSC<sub>E</sub>) as follows:

(CSC<sub>E</sub>) Necessarily, for any subject S, S is *conscious* at *t* if and only if S is conscious *of S* at *t*.

As I have emphasised, many authors understand the notion of self-consciousness (in the egological sense) in slightly stronger terms, namely as a consciousness *of oneself as oneself*. While it will be important to keep this in mind in our discussion of (CSC<sub>E</sub>), we need not build this stronger notion into the formulation of the claim.

We could formulate a *third* variant of (CSC) that draws upon the 'deflationary' notion of self-consciousness discussed in the previous section; but of course this would be a mere tautology:

(CSC<sub>DEF</sub>) Necessarily, for any subject S, S is *conscious* at *t* if and only if S is (*herself*) conscious at *t*.

I will leave (CSC<sub>DEF</sub>) aside for now, although it will come to play a dialectical role in the next chapters. Indeed, for the notions of self-consciousness at play in (CSC<sub>NE</sub>) and (CSC<sub>E</sub>) to be philosophically interesting and potentially illuminating, it should be clear that such a notion does not collapse into the deflationary notion of (CSC<sub>DEF</sub>) – namely, into the mere notion of consciousness.

The implications of (CSC<sub>NE</sub>) and (CSC<sub>E</sub>) are not immediately obvious. *Prima facie*, both claims are neither equivalent to, nor entailed by, any of the foundational claims about consciousness provided in chapter 1. This seems particularly clear in the case of (CSC<sub>E</sub>); the case of (CSC<sub>NE</sub>) is more complicated in so far as it is not immediately clear what being conscious of *one's conscious experience* really entails. In the next two chapters, I will examine each claim in turn, by looking at how contemporary proponents of these claims have explained them. Before turning to contemporary formulations of (CSC<sub>NE</sub>) or (CSC<sub>E</sub>), however, it is helpful to review their philosophical ancestry – for each of them has a surprisingly long and interesting history.

## 2.3 Historical overview

I have offered a first pass at disambiguating the idea that consciousness constitutively involves self-consciousness (CSC). I shall now provide some textual evidence for (CSC<sub>NE</sub>) and (CSC<sub>E</sub>) in the history of philosophy. The notions of non-egological self-consciousness and egological self-consciousness both have a long history in philosophy, as does the idea that one of them, or indeed both of them, are constitutively involved in consciousness in general.

### 2.3.1 Non-egological self-consciousness

#### 2.3.1.1 Ancient and medieval philosophy

The idea that consciousness constitutively involves non-egological self-consciousness, or consciousness of *consciousness itself*, can arguably be traced all the way back to the work of Aristotle.<sup>7</sup> In *Metaphysics* 12.9, Aristotle seems to argue that conscious mental states

<sup>7</sup>See Caston (2002) for a defense of this historical claim.

take *themselves* as their own intentional object, aside from whatever their primary intentional object may be:

It seems that knowing, perceiving, believing and thinking are always of something else, but of themselves on the side.

Aristotle, *Metaphysics* 12.9 (translation from Caston 2002, p. 786)

The following passages from Aristotle's *Nicomachean Ethics* and *On the Soul* are more explicit about the nature of the relation between the subject and her conscious mental states; indeed, Aristotle argues that one cannot perceive something without 'perceiving' that one perceives it:

The person seeing perceives that he is seeing, the person hearing [perceives] that he is hearing... and similarly in other cases there is something that perceives that we are in activity, so that we will perceive that we perceive and think that we think.

Aristotle, *Nicomachean Ethics* 9.9 (translation from Caston 2002, p. 774)

[When we see and hear] we perceive that we see and hear.

Aristotle, *On the Soul* 3.2 (translation from Caston 2002, p. 760)

It is worth noting that the verb rendered by 'perceive' (αἰσθάνομαι in Greek) can also be translated by 'to be aware of'. Aristotle uses a different word in the following passage from *On Perception and Perceptibles*:

It is impossible to perceive and see something and yet be unaware of perceiving it.

Aristotle, *On Perception and Perceptibles* 3.2  
(translation modified from Caston 2002, p. 758)

In the above passage, the verb translated by 'being unaware' is λανθάνειν, which literally means 'to escape notice' or 'to be unknown or unseen'. Thus, Aristotle's claim appears to be that one cannot consciously see something and yet fail to take notice of this experience. This choice of terminology raises the issue of whether Aristotle's thesis is an epistemic claim about the subject's *knowledge* of her occurrent conscious mental states, or a phenomenological claim about the subject's *consciousness* of her occurrent conscious mental states. As we shall see, this ambiguity is often present in historical claims about consciousness constitutively involving an 'awareness' of one's experience.

Interestingly, a view reminiscent of Aristotle's emerged independently in 6<sup>th</sup> century India, in the work of the Buddhist philosopher and logician Dignāga (c.480–c.540 CE).

In his magnum opus on epistemology, the *Pramāna-samuccaya* (“A Compendium of Validities”), Dignāga argued that consciousness is characterised by a kind of ‘self-awareness’ (*svasamvedana*)<sup>8</sup> in virtue of which every conscious mental state discloses itself to its subject in addition to disclosing its intentional object:

Every cognition is produced with a twofold appearance, namely, that of itself and that of the object.

Dignāga, *Pramāna-samuccaya* 1.11.ab (1968, p. 28)

This pithy passage is translated more precisely by Finnigan (2018) as follows (the words in brackets are also added by Finnigan):

[Conscious experience has] two forms [or appearances]... the cognition [or awareness] of the object and the cognition [or awareness] of that [i.e. the cognition itself].

Dignāga, *Pramāna-samuccaya* 1.11.ab (translation from Finnigan 2018, p. 3)

This passage is strikingly similar to what Aristotle writes in *Metaphysics* 12.9, and like the latter, its meaning is very debated. One matter of disagreement is whether Dignāga’s notion of ‘self-awareness’ is a perceptual or quasi-perceptual awareness of one’s conscious experience that makes a distinct contribution to the subject’s overall phenomenology, or whether it is the very fact that every conscious experience is constitutively such that it contributes to its subjects overall phenomenology.<sup>9</sup>

A few examples drawn from the secondary literature on Dignāga show that it is difficult to assess to what extent his view on ‘self-awareness’ really goes beyond foundational claims put forward in chapter 1. Thus, Arnold (2010) interprets Dignāga as claiming that one has a “first-personal acquaintance with the occurrence of one’s own cognitions [i.e., one’s own conscious mental states]” (pp. 356-7). On Coseru (2016)’s interpretation, “self-awareness consists in conscious cognitive events being inherently self-revealing” (p. 534). Finally, MacKenzie (2016) articulates Dignāga’s view as follows: “when I have an experience as of

<sup>8</sup>In Sanskrit, *samvedana* (संवेदना) can have both an epistemic sense (‘knowledge’, ‘cognition’) and a phenomenological sense (‘perception’, ‘sensation’, ‘feeling’). In turn, *svasamvedana* (स्वसंवेदन) is a lexical compound formed from the prefix *sva-* (*self-*) and *samvedana* (I am grateful to Michele Bianconi for confirming this observation).

<sup>9</sup>See Arnold (2010), Kellner (2010), Coseru (2016), and Finnigan (2018) for a discussion.

a tree, on this view, the experience presents both the tree... and the experiencing of the tree” (p. 113). It is unclear whether or not, on these interpretations of Dignāga, his theory of ‘self-awareness’ should plausibly follow from foundational claims about consciousness, or whether it is a matter of substantive theory. As we shall see in the next chapter, the same ambiguity is present in some contemporary versions of the claim that consciousness constitutively involves self-consciousness in the non-egological sense.

After Dignāga, other philosophers of the Yogācāra school of Buddhist epistemology, such as Dharmakīrti (6<sup>th</sup> or 7<sup>th</sup> century) and Śāntaraksita (725-788), endorsed the claim that consciousness constitutively involves a form of awareness of itself.<sup>10</sup> A common metaphor found in the Yogācāra school is that of ‘self-luminosity’ (*svaprakāśa*), according to which consciousness reveals *itself* as well as its objects, just like the light of a candle makes *itself* visible while also illuminating other things.<sup>11</sup>

Within the Buddhist tradition, a similar view was also defended many centuries later by Tibetan thinkers such as Gorampa (1429-1489), who wrote in his *Elucidation of Epistemology*:

With regard to the object, consciousness and the consciousness of that consciousness appear to consciousness in a twofold manner... Reflexive awareness is this twofold appearance. If there were no reflexive awareness there would be no perception.

Gorampa, *Elucidation of Epistemology* 300 (translation from Garfield 2015, p. 148)

To my knowledge, this is one of the earliest occurrences of the expression ‘consciousness of consciousness’, whose use was later revived within the phenomenological tradition. On the view articulated by Gorampa, consciousness is intrinsically self-referential: all consciousness is consciousness *of itself*, aside from being consciousness of an intentional object. Such ‘reflexive awareness’, as Gorampa calls it, is taken to be a necessary condition for there to be something it is like *at all* to be in a mental state.

<sup>10</sup>For example: “Consciousness is reflexive awareness; the nature of consciousness is reflexive awareness and that which is not reflexively aware is insentient” (Śāntaraksita, *The Ornament of the Middle Way* 16, translated by Blumenthal 2004, p. 237).

<sup>11</sup>See MacKenzie (2007) and MacKenzie (2017) on the metaphor of self-luminosity. Interestingly, a similar metaphor for self-consciousness has been independently developed within Medieval Islamic philosophy, in particular in the work of the 12<sup>th</sup> century Persian philosopher Suhrawardī (see Suhrawardī, *Hikma al-ishrāq* II.1.5.117-118, translated by Kaukua 2015, pp. 149-50).

### 2.3.1.2 Early modern philosophy

In Europe, a similar account of consciousness was developed in early modern philosophy in the wake of Cartesianism. Whether Descartes himself endorsed this account remains controversial on the basis of scarce textual evidence. Like most 17<sup>th</sup> century philosophers before Locke, Descartes did not develop an explicit theory of ‘consciousness’ as such (as the French *conscience* and the Latin *conscientia* did not yet have a fixed meaning at the time). Nonetheless, Descartes’ very broad concept of ‘thought’ or ‘mental state’ (*cogitatio*) – more or less synonymous with what we would now call *conscious experience* – does seem to involve a form of self-consciousness in the non-egological sense:<sup>12</sup>

By the word ‘thought’ [*cogitatio*], I understand all those things which occur in us while we are conscious, *insofar as the consciousness of them is in us*. And so not only understanding, willing, and imagining, but also sensing, are here the same as thinking.

Descartes (1644/1983, p. 5, my emphasis)

I use the term *thought* [*cogitatio*] to cover everything that is in us in such a way that *we are immediately conscious of it*. Thus all operations of the will, the intellect, the imagination, and the senses are thoughts. But I added ‘immediate’, so as to exclude the consequences of these operations: for instance, voluntary motion certainly has thought at its origin, but is not itself a thought.

Descartes (1641/2008, p. 102, my emphasis)

A few passages suggest that Descartes specifically understands consciousness in terms of a second-order thought (or mental state) targeting a first-order thought: “to be conscious [*conscium esse*] is both to think and to reflect upon one’s thought” (Descartes 1648/1991, p. 335, translation slightly modified). However, other passages suggest that the act of thinking (experiencing) and the act of thinking about one’s thought are in fact one and the same.<sup>13</sup> In any case, Descartes does not offer a fully developed theory of the self-referential nature of consciousness, and it may not be possible on the basis of his writings to determine exactly what kind of view he held.

The Cartesian philosopher Antoine Arnauld (1612-1694) is more explicit than Descartes about his endorsement of the idea that consciousness is ‘conscious of itself’:

<sup>12</sup>On whether Descartes endorsed the idea that consciousness involves a form of self-consciousness (in the non-egological sense), see Lähteenmäki (2007), Thiel (2011, pp. 45-8), and Strawson (2013, pp. 14-5).

<sup>13</sup>See, for example, Descartes (1641/2008, pp. 230-1).

[T]hought or perception is essentially reflective on itself, or, as it is said more aptly in Latin, *est sui conscia* [is conscious of itself].<sup>14</sup>

Arnauld (1683/2011, p. 73, my translation)

[T]he distinction between those beings who are intelligent [and conscious], and those who are not, comes from the fact that the former *sunt conscia sui, et suae operationis* [are conscious of themselves, and of their mental operations], while the latter are not.

Arnauld (1683/2011, p. 53, my translation)

While Arnauld switches from French to Latin in the original text in an attempt to be more specific, these excerpts are indicative of the difficulty of formulating (CSC<sub>NE</sub>) clearly. The second passage forebodes the tendency to equivocate between the non-egological and the egological notions of self-consciousness, given that being conscious of *one's mental states* (or 'operations') and of *oneself* appear to be treated as one and the same thing. The first passage seemingly avoids this equivocation by discarding the reference to subjects of experience altogether; but it remains unclear how to interpret the idea that *experiences* are conscious of themselves, if not as shorthand for the experiences' *subject* being conscious of them.

In the English-speaking world, Locke was one of the first philosophers to give a specific meaning to the term 'consciousness,' dissociated from the evaluative meaning of 'conscience' (while the French '*conscience*' carried both meanings). In this *Essay Concerning Human Understanding*, Locke defines consciousness as "the perception of what passes in a Man's own mind" (Locke 1689/1979, p. 115); several other passages of the *Essay* gloss this notion in a way that suggests that consciousness involves a form of self-consciousness in the non-egological sense:

[It is] altogether as intelligible to say, that a body is extended without parts, as that anything thinks without being conscious of it, or perceiving, that it does so... [T]hinking consists in being conscious that one thinks.

Locke (1689/1979, p. 115)

It [is] impossible for any one to perceive, without perceiving that he does perceive. When we see, hear, smell, taste, feel, meditate, or will any thing, we know that we do so.

Locke (1689/1979, p. 335)

<sup>14</sup>Like Descartes, Arnauld used the term 'thought' very broadly as a synonym of conscious experience.

On a common reading of these passages, Locke's theory of consciousness prefigures what we would now call the 'Higher-Order Perception' (HOP) theory of consciousness.<sup>15</sup> According to the HOP theory, prominently defended by Armstrong (1968) and Lycan (1996), mental states are conscious because we 'perceive' them via a kind of *inner sense*, namely a high-order perceptual (or quasi-perceptual) mental state that takes the first-order mental state as its object. Leibniz already ascribed such a theory to Locke, and criticised it on the grounds that it would entail an infinite regress: if, to perceive a yellow lemon, I must perceive my first-order perception of the yellow lemon, then I must also perceive this second-order perception by means of a *third-order* perception, and so on *ad infinitum*.<sup>16</sup>

Modern defenders of the HOP theory can escape such regress by claiming that the second-order mental state in virtue of which the first-order mental state is *conscious* is not itself a *conscious* mental state; consequently, the second-order mental state need not be the object of yet another (third-order) mental state, and there is no threat of infinite regress. This line of defence was not available to Locke, for, like many other early modern thinkers, he did endorse the claim that all mental states (all thoughts, perceptions, or operations of the mind) are *conscious* mental states. However, many modern commentators have suggested that Locke's theory of consciousness is not actually a HOP theory.<sup>17</sup> On this interpretation, Locke's notion of consciousness "is not an additional mental act above and beyond the original perception" (Weinberg 2016, p. 33). Rather, perceiving a yellow lemon constitutively involves being conscious of one's perception of the yellow lemon, not by means of a higher-order mental state targeting one's first-order perception of the yellow lemon, but through a "reflexive self-referential awareness internal to [the first-order] ordinary perception" (Weinberg 2016, p. 32). This notion of a "reflexive self-referential awareness" stands in need of further elucidation. Unfortunately, it seems difficult to find a technical account of such a notion in Locke's work. Like Arnauld, Locke did not fully

<sup>15</sup>See, among many other examples of this reading of Locke, Lycan (1995) and Carruthers (2004).

<sup>16</sup>See Leibniz (1704/1996, p. 118).

<sup>17</sup>See, for example, Coventry and Kriegel (2008), Thiel (2011, pp. 109-120), and especially Weinberg (2016).

explain how we should understand the kind of reflexivity that is, on his view, inherent to consciousness.

### 2.3.1.3 Post-Kantian philosophy

In the 19<sup>th</sup> century, Franz Brentano (1838-1917) revived and refined the idea that consciousness essentially involves a form of consciousness of one's experience, which he called 'inner consciousness'. According to Brentano, every conscious mental event has two objects, one of which is itself: all experiences are in some way self-directed, aside from being directed at an intentional object. When one hears a musical note, for example, the note is the *primary* object of one's hearing, while one's hearing of the note is the *secondary* object of one's hearing:<sup>18</sup>

Every mental act is conscious; it includes within it a consciousness of itself. Therefore, every mental act, no matter how simple, has a double object, a primary and a secondary object. The simplest act, for example the act of hearing, has as its primary object the sound, and for its secondary object, itself, the mental phenomenon in which the sound is heard.

Brentano (1874/1995, p. 119)

As many commentators have remarked, these passages suggest that Brentano is a forerunner of a contemporary theory of consciousness known as *self-representationalism*, according to which a mental state is conscious if and only if it represents itself in the right way.<sup>19</sup> In the above passages, Brentano does not formulate his view in terms of representation, but in terms of *presentation* (*Vorstellung*): every (conscious) mental state *presents* an intentional object, but also presents *this very presentation*; every (conscious) mental state is, in this sense, *self-presenting*.<sup>20</sup> It is clear from other passages that this inner presentation of the mental state to its subject is a form of 'inner perception' (*innere Wahrnehmung*).

Like Locke, Brentano emphasised that this inner perception is not a form of inner observation (*innere Beobachtung*) – it is not a matter of attending to, or reflecting upon,

<sup>18</sup>See also Brentano (1874/1995, p. 98).

<sup>19</sup>See Kriegel (2009) for an exposition and defence of self-representationalism, and Kriegel (2013) for a discussion of Brentano's thesis in light of self-representationalism. I will come back to self-representationalism in chapter 3.

<sup>20</sup>Brentano uses the term 'self-presentation' (*Selbstvorstellung*).

one's mental states (Brentano 1874/1995, p. 22). When we experience raging anger – to take Brentano's own example – we do not experience it by means of introspection or inner observation, for this would paradoxically require us to calm down in order to focus on our present experience. Unlike introspection or inner observation, inner perception of one's experience is neither mediated by attention, nor under voluntary control. One does not *decide* to perceive one's experience by attending to it; rather, every experience is 'self-presenting' in the very same process in which it presents its primary object.

Brentano's treatment of the 'self-presenting' nature of consciousness was quite influential in the 20<sup>th</sup> century, especially within the phenomenological tradition. The founder of this tradition, Edmund Husserl (1859-1938), was himself a student of Brentano. Although Husserl did criticise some aspects of Brentano's thesis about inner perception, he endorsed the gist of the claim, as is apparent in the following excerpts:

Every act [i.e. conscious event] is consciousness of something, but there is also consciousness of every act. Every experience is 'sensed', is immediately 'perceived' (internal consciousness).

Husserl (1905/1991b, p. 130)

[E]very experience is 'consciousness' and consciousness is consciousness *of*... But every experience is *itself experienced* [erlebt], and *to that extent* also 'conscious' [bewußt]. This being-conscious [Bewußt-sein] is consciousness of the experience...

Husserl (1907/1991a, p. 301, translation slightly modified)

This consciousness is self-perceiving... This perceiving that presents all lived-experiencing to consciousness is the so-called inner consciousness or inner perceiving.

(Husserl 1920/1970, p. 320; translation from Zahavi 2005, p. 130)

Husserl argues that conscious experience is characterised by a certain kind of 'self-manifestation' or 'self-appearance' (*Für-sich-selbst-erscheinens*; Husserl 1924/1959, p. 189). Thus, according to him, "self-consciousness, rather than being something that only occurs during exceptional circumstances... is a feature characterizing the experiential dimension as such" (Zahavi 2014, p. 12). The relevant notion of self-consciousness is non-egological: it is form of consciousness *of one's experience*.

Unlike Brentano, however, Husserl cautions against an understanding of this non-egological self-consciousness by analogy with our consciousness of external objects. For Brentano, as we have seen, every conscious mental state has a primary (external) object

and itself as a secondary object, and ‘inner perception’ consists in the mental state having itself as a secondary object. For Husserl, by contrast, the ‘self-appearance’ of experience should *not* be understood on an act-object model, because our experiences are not presented to us as external objects are presented to us *through them*. Indeed, on Husserl’s view, consciousness of external objects involves a subject-object division – a division between *what* is presented, and the subject *to whom* it is presented. The ‘self-appearance’ of consciousness exhibits no such subject-object division. This is why Husserl often avoids the indirect construction ‘consciousness of one’s experience’, and writes instead that experience is ‘experienced’ or ‘lived through’ (*erlebt*). As he puts it, that something “is experienced, and is in this sense conscious, does not and cannot mean that this is the *object* of an act of consciousness, in the sense that a perception, a presentation or a judgement is directed upon it” (Husserl 1901/2001b, p. 273). It is not obvious that such formulations point to an aspect of phenomenology that is not also acknowledged by the Nagelian account of consciousness, in so far as what it is for a mental state to be ‘experienced’, ‘lived through’ or ‘self-manifesting’ simply is for that mental state to constitutively contribute to its subject’s overall phenomenology.

After Husserl, the phenomenologist Jean-Paul Sartre (1905-1980) explicitly endorsed the idea that consciousness constitutively involves ‘consciousness of consciousness’:

[T]he type of existence that consciousness has is that it is consciousness of itself.

Sartre (1936/2004, p. 4)

[T]he necessary and sufficient condition for a knowing consciousness to be knowledge of its object, is that it be consciousness of itself as being that knowledge. This is a necessary condition, for if my consciousness were not consciousness of being consciousness of the table... it would be a consciousness ignorant of itself, an unconscious consciousness [*conscience inconsciente*] – which is absurd. This is a sufficient condition, for my being conscious of being conscious of that table suffices in fact for me to be conscious of it... What is this consciousness of consciousness?... This self-consciousness [is] *the only mode of existence which is possible for a consciousness of something*.

Sartre (1943/1948, pp. lii-liv, translation slightly modified)

Sartre is particularly clear that the relevant notion of self-consciousness is *not* a form of consciousness of *oneself*, that is, an egological notion of self-consciousness (Sartre 1936/2004, p. 8). For him, although I am not aware of *myself* at all times

when I am conscious, I cannot be conscious of anything without being conscious of that very consciousness. Importantly, Sartre also stresses that “this consciousness of consciousness... is not *positional*, i.e. consciousness is not its own object” (Sartre 1936/2004, p. 5). In Sartre’s vocabulary, consciousness is said to be ‘positional’ if it is directed towards an intentional object which is presented (or ‘posited’) to the subject as ‘transcendent’, i.e. distinct from the subject’s experience. The kind of self-consciousness inherent to every conscious experience is not ‘positional’, because I am not conscious of my conscious experiences in the same way in which I am aware of intentional objects – as Husserl emphasised before Sartre.

Continuing a common trend in the history of philosophy, Sartre emphasises that this form of self-consciousness is not a matter of *reflecting* upon one’s experience (Sartre 1943/1948, pp. liii-liv). Rather, it is ‘pre-reflective’ and constitutive of the first-order conscious experience: “[t]his self-consciousness we ought to consider not as a new consciousness, but as *the only mode of existence which is possible for a consciousness of something*” (Sartre 1943/1948, p. liv).

A few years after the publication of Sartre’s *Being and Nothingness*, the American phenomenologist Aron Gurwitsch also defended the claim that consciousness essentially involves ‘self-awareness’ or ‘inner awareness’ in a manuscript titled *Marginal Consciousness*. In the first chapter, ‘The Self-Awareness of Consciousness’, Gurwitsch presents his view as a development of Brentano’s theory of consciousness. According to him, self-awareness is a necessary condition for consciousness:

When an object is given in experience, the experiencing subject is conscious of the object and has an awareness of this very consciousness of the object... Thus every act of consciousness is accompanied by an awareness of itself.... *That every act of consciousness carries self-awareness with it, so that this self-awareness accompanies us throughout all our conscious life, is more than a merely empirical fact, ascertained with utmost empirical generality; it is an a priori condition for consciousness...* [S]elf-awareness permanently and necessarily pervades all of our conscious life, so that at every moment of this life we are aware of the act experienced at this very moment.

Gurwitsch (1953/2010, pp. 451-5)

Gurwitsch also insists that self-awareness is ‘unreflective’ or ‘marginal’: “the experience of every act, whatever its object, carries marginal consciousness of... the act in question”

(Gurwitsch 1953/2010, p. 463).<sup>21</sup> In other words, Gurwitsch reiterates the point that the relevant form of self-awareness does not require directing one's attention to one's recurrent experience.

Thus, the idea that consciousness constitutively involves a non-egological and 'pre-reflective' form of self-consciousness is a theme that runs throughout the history of the phenomenological tradition, from Brentano to Gurwitsch through Husserl and Sartre.<sup>22</sup> This view is certainly not as common within 20<sup>th</sup> century analytic philosophy. Nonetheless, one occasionally finds formulations that seem reminiscent of the phenomenological view. For example, Alvin Goldman argues that thinking always involve a non-reflective meta-awareness that one thinks, in a passage often quoted as a rare instance of this view in analytic philosophy:

In the process of thinking about *x*, there is already an implicit awareness that one is thinking about *x*. When we are thinking about *x*, the mind is focused on *x*, not on our thinking of *x*. Nevertheless, the process of thinking about *x* carries with it a non-reflective self-awareness.

Goldman (1970, p. 96)

While Goldman focuses on *thinking* rather than *experiencing*, many have read him as defending the claim that consciousness constitutively involves a form of self-consciousness.<sup>23</sup> Another passage often quoted as an 'early' expression of the view that non-egological self-consciousness is constitutive of consciousness in the analytic tradition is due to Harry Frankfurt:

[B]eing conscious in the everyday sense does (unlike unconsciousness) entail reflexivity: it necessarily involves a secondary awareness of a primary response... For what would it be like to be conscious of something without being aware of this consciousness? It would mean having an experience with no awareness whatever of its occurrence. This would be, precisely, a case of unconscious experience. It appears, then,

<sup>21</sup>Here, 'act' means 'mental act', which refers in the phenomenological tradition to intentional mental states.

<sup>22</sup>Dan Zahavi goes so far as to write that "literally all the major figures in phenomenology defend the view that the experiential dimension is characterized by a tacit self-consciousness" (Zahavi 2005, p. 11; see also Zahavi 2014, p. 30).

<sup>23</sup>See, among many other examples, Kriegel (2009, p. 176), S. Gallagher and Zahavi (2012, p. 51), and Strawson (2013, p. 13). To my knowledge, Goldman never provided a retrospective commentary on this passage to clarify what he intended to say.

that being conscious is identical with being self-conscious. Consciousness is self-consciousness.... The self-consciousness in question is a sort of *immanent reflexivity* by virtue of which every instance of being conscious grasps not only that of which it is an awareness but also the awareness of it. It is like a source of light which, in addition to illuminating whatever other things fall within its scope, renders itself visible as well.

Frankfurt (1988, pp. 161-2)

The view expressed by Frankfurt in this passage is remarkably similar to older views discussed in this section. Thus, the idea that an experience whose subject was not *conscious of* it would be an *unconscious* experience (which is an inconsistent notion) is already found in Sartre.<sup>24</sup> Likewise, Frankfurt's notion of 'immanent reflexivity' is reminiscent of Śāntaraksita and Gorompa's ideas, as is the metaphor of light which is very similar to the Buddhist notion of 'self-luminosity'. Frankfurt also clarifies in a footnote that the sort of self-consciousness he has in mind is not egological and does not refer to a propositional attitude:

What I am here referring to as 'self-consciousness' is neither consciousness of a self – a subject or ego – nor consciousness that there is awareness. Both require rational capacities beyond what would seem to be necessary for consciousness itself to occur. The reflexivity in question is merely consciousness's awareness of itself. To hear a sound consciously, rather than to respond to it unconsciously, involves being aware of hearing it or being aware of the sound as heard.

Frankfurt (1988, pp. 162, fn. 5)

From these passages, it is unclear whether Frankfurt's notion of self-consciousness really goes beyond the foundational claims outlined in chapter 1. Indeed, his purpose seems to be to highlight that conscious experiences are not merely events that happen to us; they also constitutively contribute to what it is like for us. If being conscious or aware of one's conscious experience merely points to the fact that conscious experiences contribute to one's overall phenomenology, then (CSC<sub>NE</sub>) plausibly follows from the Nagelian elucidatory account of consciousness. I will come back to this issue in the next chapter, in which I will discuss in more detail contemporary formulations of this claim. As we shall see, one version of the claim (inspired by Sartre and Husserl) does not seem to go

<sup>24</sup>See also Locke (1689/1979, p. 110): "If the *Soul* doth *think in a sleeping Man*, without being conscious of it, I ask, whether, during such thinking, it has any Pleasure or Pain, or be capable of Happiness or Misery?... For to be happy or miserable without being conscious of it, seems to me utterly inconsistent and impossible".

beyond foundational claims about consciousness, while another (drawing upon Brentano and Gurwitsch) does seem to point to a distinct substantive view of consciousness.

## 2.3.2 Egological self-consciousness

### 2.3.2.1 Ancient and medieval philosophy

The claim that consciousness constitutively involves self-consciousness in the *egological sense* – consciousness *of oneself (as oneself)* – appears to be less prevalent in the history of philosophy than its non-egological counterpart. In Ancient philosophy, Stoics such as Hierocles and Seneca might have endorsed the view that all animals, including humans, are conscious of themselves from birth:

We should realise that when an animal is born it immediately simultaneously perceives itself.

(Hierocles, translation from Knuuttila and Sihvola 2014, p. 417)

[A child] does not know what an animal is, but he is aware that he is an animal... Thus, even children and animals have an awareness of their primary constituent, but the awareness is not very lucid, nor articulate.

(Seneca, *Letter 121*, translation from Knuuttila and Sihvola 2014, pp. 417-8)

In Late Antiquity, Augustine (354-430 CE) examined the way in which the mind may relate to itself, although he did not clearly distinguish between a phenomenological and an epistemic notion of self-awareness (i.e. what we may call a sense of self on the one hand, and self-knowledge on the other). In his treatise *On the Trinity*, he argues that there are two levels of self-awareness: an implicit, non-conceptual self-awareness that even infants have (*nosse*), and a more intellectual form of self-awareness that requires the ability to think about oneself (*cogitare*):

Should we believe that [the infant's mind] is aware of [*nosse*] itself, but, being too intent on the things that it has begun to sense with the bodily senses, with pleasure all the greater for being new, while it cannot be unaware of [*ignorare*] itself, is yet not able to think of [*cogitare*] itself? [Infants] have no thought of their inner self, nor can they be admonished to do so... But we have already shown in this same book that it is one thing not to be aware of [*nosse*] oneself, and another thing not to think of [*cogitare*] oneself.

Augustine (2002, pp. 143-4)

Many centuries after Augustine, the Persian philosopher Ibn Sīnā (c. 980-1037), also known as Avicenna in the Christian world, developed a more sophisticated theory of self-consciousness. In a very influential passage from his psychological treatise *On the Soul* which anticipates the Cartesian *cogito*, Avicenna outlines a thought experiment known as the ‘Flying Man’, in which he asks the reader to imagine what it would be like to be deprived of any sensory input:

[O]ne of us must imagine himself so that he is created all at once and perfect but his sight is veiled from seeing external [things], that he is created floating in the air or in a void... He will not hesitate in affirming that his self exists, but he will not thereby affirm any of his limbs, any of his intestines, the heart or the brain, or any external thing... Thus, he who takes heed has the means to take heed of the existence of the soul... and to know and be aware of it.

(Avicenna, *On the Soul* 5.7, translation from Kaukua 2015, p. 35)

The interpretation of this thought experiment and of Avicenna’s commentary is debated. What seems relatively uncontroversial is that he wants to establish the distinction between the soul and the body, by arguing that one could know that one exists even in the absence of sensory and bodily stimuli. However, commentators have debated whether or not such knowledge is supposed to be justified by a non-sensory form of self-consciousness or self-awareness which is also present in ordinary waking life. In a recent monograph dedicated to the notion of self-awareness in Islamic philosophy, Jari Kaukua argues that the Flying Man is intended as a demonstration that self-awareness accompanies all conscious states: Avicenna “takes self-awareness to be a phenomenal feature of experience, not a mere transcendental or logical condition” (Kaukua 2015, p. 36), and wants to convey that it is “something familiar to each of us from a perfectly commonplace human experience” (p. 85). According to this interpretation, the Flying Man is meant to make us aware of a feature of experience that usually goes unnoticed. As Avicenna himself puts it, “it is often the case that knowledge about something is close at hand, but one does not pay attention to it, so that it verges on the unknown and is investigated at the greatest remove”.<sup>25</sup> Beyond Avicenna, many other Islamic philosophers commented on the Flying

<sup>25</sup> Avicenna, *On the Soul* 5.7, translation from Kaukua 2015, p. 86.

Man and appear to have reached similar conclusions about the ubiquity of a basic form of self-awareness in experience<sup>26</sup>.

### 2.3.2.2 Early modern philosophy

In early modern philosophy, the notion of self-consciousness remains linked to the notion of self-knowledge in the wake of Descartes' *cogito* argument, being typically defined as that in virtue of which one knows that one exists. Interestingly, some of the philosophers who hold that consciousness constitutively involves self-consciousness *in the non-egological sense* also appear to hold that it constitutively involves self-consciousness *in the egological sense*. For example, Locke does not merely argue that perceiving requires perceiving one's perception; he also seems to argue that perceiving involves being conscious *of oneself*, and consequently of one's own existence:

In every act of sensation, reasoning, or thinking, we are conscious to our selves of our own being.

Locke (1689/1979, p. 619)

[O]ur own existence is known to us by a certainty yet higher than our senses can give us of the existence of other things, and that is internal perception, a self-consciousness or intuition.

(Locke, journal note from 1696 printed in King 1830, p. 138)

These passages might simply be read as Locke's version of Descartes' *cogito*: in having any conscious experience, one is *in a position to know* that one exists. Being in a position to know that one exists need not involve being actually *conscious of oneself (as oneself)*. However, one can also read these passages as making the claim that all conscious experiences involve some a sense of self – a sense that *I myself* exist. As Locke also puts it, “consciousness [is that] whereby I am my *self* to my *self*” (1979, p. 345). In a rich monograph on early modern theories of consciousness and self-consciousness, Udo Thiel notes that “the consciousness of states and operations on the one hand and the consciousness of one's own existence or self-consciousness on the other are very closely connected in Locke” (2011, p. 118). On this interpretation of Locke, “by way of the consciousness we have of mental states

<sup>26</sup>These philosophers include Suhrawardī (1154-1191) and, much later, Mullā Sadrā (c. 1571-1640); see Kaukua (2015) for a detailed discussion.

and operations, we perceive immediately... our own existence” (*ibid.*, p. 119). Ultimately, it is difficult to work out a specific theory of self-consciousness (in the egological sense) from Locke’s work, as he remains somewhat evasive on how one’s perception of one’s mental states relates to one’s perception of oneself – or indeed, whether one’s perception of one’s mental states only puts one *in a position to know* about one’s own existence.

After Locke, the German philosopher Christian Wolff (1679-1754) clearly claims that consciousness constitutively involves being conscious of *oneself*. Indeed, Wolff argues, in order to be conscious of anything in particular, the subject has to be able to *distinguish* intentional objects from one another, otherwise there would only be a chaotic bundle of ideas in experience and conscious mental states would not be about anything specifically. But in distinguishing intentional objects from one another, he claims that we necessarily become conscious of our own activity of distinguishing, hence of ourselves. This is what leads him to conclude that being self-conscious (in the egological sense) is necessary to be conscious of anything:

[In being conscious of things] we are conscious of ourselves, namely, we notice the difference between ourselves and other things of which we are conscious.... Since, supposing that we are conscious of something which we cognize through the sense, then we must notice the distinction between what we perceived in it, indeed, we must immediately distinguish the thing that we cognize thereby from the other things. Yet representing things is an action of the soul no less than this differentiation... and we cognize thereby the distinction of the soul from the things that it represents and that it distinguishes.

Wolff, *German Metaphysics*, §§729-30 (translation from Dyck 2014, p. 107)

Let us unpack this difficult passage. According to Wolff, the ‘soul’ (i.e., the subject) can be only conscious of an object (e.g., a yellow lemon) by discriminating it from other objects (for example, other objects in the visual scene) – otherwise we would not be conscious of anything in particular, and perception would just be a chaotic assemblage of sensations. In discriminating objects from one another, the soul also becomes conscious of its own activity, which, in turn, means that it becomes conscious of itself as distinct from the perceived objects. Since being conscious of objects necessarily involves a mental act of discriminating between objects, and since a such mental act necessary involves discriminating between the

objects and the subject herself, it follows on Wolff's view that being conscious of objects necessarily involves being conscious of oneself.<sup>27</sup>

Immanuel Kant (1724–1804), who was heavily influenced by Wolff, appears to endorse a weaker claim, according to which having any experience merely requires the *capacity* to self-ascribe that experience: “The ‘I think’ must be able to accompany all my representations” (Kant 1781/1999, p. 246). Kant's account of self-reference and self-knowledge is extremely complex, and I cannot discuss it in any detail here. However, it is fairly clear that Kant did not intend his emphasis on the role of the ‘I think’ to entail that all conscious experiences involve a form of consciousness *of oneself*.<sup>28</sup>

### 2.3.2.3 Post-Kantian philosophy

The idea that consciousness *of oneself* permeates our mental lives became quite popular in the early days of psychology. Thus, William James argues that there is always an elusive sense of self in the background of experience: “whatever I may be thinking of, I am always at the same time more or less aware of *myself*, of my *personal existence*” (James 1892/1961, p. 42; see also Calkins 1908, p. 68). In turn, Sigmund Freud seems to endorse a similar claim at least with respect to *ordinary* experience, and suggests that this sense of self, far from being elusive, is rather manifest: “normally, there is nothing of which we are more certain than the feeling of our self, of our own ego” (Freud 1930/1961, p. 12).

Some passages from Husserl suggest he held the view that subjects of experience are necessarily conscious of themselves:<sup>29</sup>

The consciousness in which I am conscious of myself is my consciousness, and my consciousness of myself and I myself are concretely considered identical. To be a subject is to be in the mode of being aware of oneself.

(Husserl 1922/1973b, p. 151; translation from Zahavi 2014, p. 13)

I exist for myself and am continually given to myself, by experiential evidence, as ‘*I myself*’.

<sup>27</sup>See Thiel (2011, pp. 304–11) for a clear explanation of Wolff's view on the interdependence of consciousness and self-consciousness.

<sup>28</sup>See Longuenesse (2017) for a detailed discussion of Kant's account of self-reference and self-knowledge. Longuenesse mentions that on Kant's view, “sensations... are not, unless they are taken up in intuition and thought under concepts, accompanied with *self-consciousness*” (p. 199, n. 37).

<sup>29</sup>See also Husserl (1932/1973a, pp. 492–3).

Husserl (1931/1960, p. 68)

In reading these passages, however, one should be careful not to ascribe to Husserl an inflationary view about the ubiquity of self-consciousness in the egological sense. Indeed, the notion of selfhood at work in these passages is very minimal, in so far as Husserl often equates the self with the stream of consciousness itself. In his *Bernau Manuscripts on Time Consciousness*, for example, he writes that the “a stream of consciousness... is necessarily conscious of itself” (Husserl 1918/2001a, p. 48, my translation), and that this is a form of ‘self-consciousness’ (*Seiner-selbst-bewusst-Sein*). There does not seem to be any daylight, on this view, between the idea that consciousness is experienced (*Erlebt*) or ‘self-manifesting’, and the idea that it is a form of consciousness *of the ‘self’*, in the specific sense that the stream of consciousness manifests *itself to itself*, as it were.

## 2.4 Conclusion

In this chapter, I offered a first pass at clarifying the claim that consciousness constitutively involves self-consciousness. As we have seen, the concept of self-consciousness has two broad meanings that are not equivalent, namely (a) consciousness *of consciousness itself* (i.e., self-consciousness in the *non-egological* sense), and (b) consciousness *of oneself* (as *oneself*) (i.e., self-consciousness in the *egological* sense).

There are two main variants of the claim that consciousness constitutively involves self-consciousness, depending on which concept of self-consciousness features in this claim. According to (CSC<sub>NE</sub>), consciousness constitutively involves consciousness *of consciousness itself*, that is, self-consciousness in the *non-egological* sense. I have suggested that (CSC<sub>NE</sub>) can more specifically formulated as the claim that a mental state is a *conscious* mental state if and only if its subject is *conscious of it*. According to (CSC<sub>E</sub>), consciousness constitutively involves consciousness *of oneself*, that is, self-consciousness in the *egological* sense. On a broad understanding of this claim, a subject is conscious if and only if she is conscious *of herself*. On a slightly stronger understanding of this claim, a subject is conscious if and only if she is conscious *of herself as herself*.

As I hope to have shown in my historical overview, both (CSC<sub>NE</sub>) and (CSC<sub>E</sub>) have a long history, arguably dating back to Ancient philosophy, and encompassing both Eastern

and Western traditions. The discussion of the history of each claim has already encountered interpretative issues that will come back in the discussion of more recent formulations. In the following chapters, I shall turn to contemporary defences of  $(CSC_{NE})$  and  $(CSC_E)$ .

### 3

## Debunking the Myth: Non-Egological Self-Consciousness

Our language may suggest that pains are perceived, but it does not suggest – and it seems to me clearly not to be true – that one perceives the feeling or the ‘having’ of one’s pains.

Shoemaker (1968, p. 564)

In the preceding chapters, I have introduced and discussed a number of foundational claims that plausibly follow from a philosophical elucidatory account of consciousness, and I have subsequently distinguished two broad notions of self-consciousness, each of which yields a distinct variant of the claim that self-consciousness is constitutive of consciousness (CSC). In this chapter, I shall focus on the first variant (CSC<sub>NE</sub>) according to which consciousness constitutively involves self-consciousness in the *non-egological sense*:

(CSC<sub>NE</sub>) Necessarily, for any subject S and any mental state M of S, M is a *conscious* mental state of S at *t* if and only if S is conscious of M at *t*.

My overall strategy in this chapter will be to argue that while various formulations of (CSC<sub>NE</sub>) can be interpreted in different ways, those that go beyond the foundational claims of chapter 1 are not adequately supported. In §3.1, I examine how several proponents of (CSC<sub>NE</sub>) gloss the Nagelian dictum in potentially distorting ways to arrive at the claim that consciousness constitutively involves consciousness of *one’s experience*. I subsequently argue that formulations of (CSC<sub>NE</sub>) fall into two categories. In the first category are claims that point to aspects of phenomenology already acknowledged by the foundational claims about consciousness outlined in chapter 1. I examine the claims that fall within this first category in §3.2, focusing on the Zahavi’s influential account of (CSC<sub>NE</sub>). In

the second category are claims that point to *determinate* aspects of phenomenology that are not acknowledged by the foundational claims as being constitutive aspects of consciousness. I examine the claims that fall within this second category in §3.3, focusing on Kriegel's equally influential account of (CSC<sub>NE</sub>). I argue that such claims are not adequately supported as substantive claims about all conscious mental states. The upshot of this analysis is that we need not appeal to self-consciousness – in the non-egological sense – to elucidate the notion of consciousness.

### 3.1 From subjectivity to consciousness of *one's experience*

Contemporary proponents of (CSC<sub>NE</sub>) often take the Nagelian dictum as a starting point to introduce the idea that consciousness constitutively involves self-consciousness (in the non-egological sense), by way of a series of emphases leading to the coinage of new – and potentially misleading – terminology.

Recall that in chapter 1, I characterised what is common to *all* conscious mental states, *qua* conscious mental states, as '*phenomenality*' – the second-order property of there being *something* (rather than *nothing*) that it is like to be in a mental state. I contrasted this notion of phenomenality with the notion of *phenomenal character*, defined as the first-order property of *what* it is like to be in some specific mental state; or, equivalently, *what* that mental state's constitutive contribution to the subject's overall phenomenology is. Furthermore, I cautioned against potential misinterpretations of the distinction between phenomenality and phenomenal character; in particular, I emphasised that it would be double-counting to consider phenomenality as a further *first-order* property alongside the phenomenal character of a mental state, for the former merely involves existential quantification over the latter.

In his seminal article, Nagel uses a slightly different terminology. After spelling out his famous dictum – an organism has conscious mental states if and only if there is something it is like *for* that organism to be *be* that organism –, he adds “[w]e may call *this* the subjective character of experience” (1974, p. 436, my emphasis). It is not entirely clear, at first, what ‘this’ refers to in this sentence. However, it quickly becomes apparent that Nagel uses the notion of ‘subjective character’ to refer to what I call ‘phenomenal character’ – following

a usage of the term now common in philosophy of mind (e.g., Shoemaker 1994a). For example, Nagel writes that “[t]he subjective character of an experience... is the essential property... in virtue of which it is, necessarily, the experience it is” (1974, p. 445, fn. 11). Commenting on experiences such as “pain, fear, hunger, and lust”, he mentions that “these experiences also have in each case a specific subjective character” (1974, p. 439). Thus, ‘subjective character’, in Nagel’s sense, qualifies *what* having a particular experience (being in a particular conscious mental state) is like for its subject.

In recent years, the terms ‘subjectivity’ and ‘subjective character’ started being used in a different way. An early example of this tendency can be found in McGinn (1991):

Subjective aspects of experience involve reference to the subject undergoing the experience – this is what their *subjectivity* consists in... Thus perceptual experiences are Janus-faced: they point outward to the external world but they also present a subjective face to their subject; they are of something other than the subject and they are like something for the subject.

McGinn (1991, p. 29)

In an often quoted passage, Levine (2001) defines the notion of ‘subjectivity’ more explicitly, alongside that of ‘qualitative character’:

Let’s take my current visual experience as I gaze upon my red diskette case, lying by my side on the computer table. I am having an experience with a complex qualitative character, one component of which is the color I perceive. Let’s dub this aspect of my experience its ‘reddish’ character. There are two important dimensions to my having this reddish experience. First... there is something it’s like for me to have this experience. Not only is it a matter of some state (my experience) having some feature (being reddish) but, being an experience, its being reddish is ‘for me’, a way it’s like *for me*, in a way that being red is like nothing for – in fact is not in any way ‘for’ – my diskette case. Let’s call this the *subjectivity* of conscious experience... The second important *dimension* of experience that requires explanation is qualitative character itself. Subjectivity is the phenomenon of there being something it’s like *for me* to see the red diskette case. Qualitative character concerns the ‘*what*’ it’s like for me: reddish or greenish, painful and pleasurable, and the like.

Levine (2001, p. 7)

What Levine calls ‘qualitative character’ is what I called ‘phenomenal character’: namely, the specific contribution that some experience makes to its subject’s overall phenomenology. In turn, what he calls ‘subjectivity’ is what I called ‘phenomenality’. Notice that Levine puts a particular emphasis, in his definition of subjectivity, on the ‘for me’ phrase.

As we have seen, both the SIL and the WIL phrases can be completed with ‘for<sup>EXP</sup> *x*’ to specify *whose* overall phenomenology an experience constitutively contributes to. For any utterance of the SIL phrase and the WIL phrase, one can meaningfully ask ‘*for whom?*’ – and the answer will have to specify a subject of experience.

Thus, Levine’s emphasis on ‘*for me*’ does not properly capture the distinction between phenomenality (there being something, rather than nothing, that it is like *for me*) and phenomenal character (what, specifically, it is like *for me*). Properly speaking, the so-called subjectivity of a conscious experience is no more *subjective* (‘*for a subject*’) and no less *qualitative* (‘*something it is like*’) than its so-called qualitative character. Consequently, if Levine’s definition of ‘subjectivity’ is intended to refer to what I call phenomenality, the emphasis should be on the word ‘something’ instead of the phrase ‘for me’, because the contrast captured by phenomenality is between there being *something* that it is like (for a subject) and there being *nothing* that it is like (for that subject).

The significance of Levine’s emphasis on the phrase ‘for me’ is amplified in some subsequent discussions of the Nagelian dictum, in which the notion of subjectivity becomes ‘subjective character’ (against Nagel’s own use of the term) and even ‘*for-me-ness*’. Here is how Kriegel (2009) introduces these notions:

When I have a conscious experience of the blue sky, there is something it is like for me to have the experience. In particular, there is a bluish way it is like for me to have it. This ‘bluish way it is like for me’ constitutes the phenomenal character of my experience. Phenomenal character is the property that makes a phenomenally conscious state (i) the phenomenally conscious state it is and (ii) a phenomenally conscious state at all. The bluish way it is like for me has two distinguishable components: (i) the *bluish* component and (ii) the *for-me* component. I call the former *qualitative character* and the latter *subjective character*. To a first approximation, phenomenal character is just the compresence of qualitative character and subjective character. To a second approximation, there is a more specific division of conceptual labor between qualitative and subjective character: a phenomenally conscious state’s qualitative character is what makes it the phenomenally conscious state it is, while its subjective character is what makes it a phenomenally conscious state at all. Thus, my conscious experience of the blue sky is the conscious experience it is in virtue of its bluishness, but it is a conscious experience at all in virtue of its for-me-ness.

Kriegel (2009, p. 1)

Kriegel’s definition of the terms ‘phenomenal character’, ‘qualitative character’ and ‘subjective character’ implicitly relies on two kinds of distinction: (a) a *determinable-determinate* distinction (b) a *type-token* distinction. On the face of it, what Kriegel calls

‘subjective character’ is what I call ‘phenomenality’: it is the second-order property that involves existential quantification over what it is like for a subject to be in a mental state. As we have seen in chapter 1, we can also gloss the relation between phenomenality and phenomenal character as a *determinable-determinate* relation – the phenomenality of an experience being the determinable of which its phenomenal character is a determinate. In turn, Kriegel’s distinction between ‘phenomenal character’ and ‘qualitative character’ appears to be a *type-token* distinction. What it is like for any subject to have a certain *type* of experience of the blue sky is, in Kriegel’s terminology, the ‘phenomenal character’ of that type of experience. By contrast, what it is like *for me* (or for some other specific subject) to have a *token* experience of the blue sky is, in Kriegel’s terminology, the ‘qualitative character’ of that token experience. Note that Kriegel does not draw a corresponding *type-token* distinction for his notion of ‘subjective character’ or ‘for-me-ness’. For the sake of completeness, we can also draw such a distinction to discriminate between there being something it is like *for any subject* to be in a mental state of a certain *type*, and there being something it is like *for a particular subject* to be in a particular *token* mental state. In summary, the type-token distinction and the determinable-determinate distinction yield four notions:

1. **Phenomenal character [determinate *type*]**

The (*type* of) contribution that a type of mental state M constitutively makes to the overall phenomenology of *any subject* who is in M.

2. **Phenomenal character [determinate *token*]**

The (*token*) contribution that a token mental state M constitutively makes to the overall phenomenology of *a particular subject* who is in M.

3. **Phenomenality [determinable *type*]**

There being some (*type* of) contribution that a type of mental state M constitutively makes to the overall phenomenology of *any subject* who is in M.

4. **Phenomenality [determinable *token*]**

There being some (*token*) contribution that a token mental state M constitutively makes to the overall phenomenology of *a particular subject* who is in M.

On this taxonomy, Kriegel's notion of 'phenomenal character' corresponds to the *first* notion above (type phenomenal character); his notion of 'qualitative character' corresponds to the *second* notion above (token phenomenal character); and his notion of 'subjective character' arguably corresponds to the *fourth* notion (token phenomenality). In light of this analysis, we can see how Kriegel's suggestion that 'qualitative character' and 'subjective character' are two *components* of 'phenomenal character' is, at the very least, misleading; with this choice of terminology occurs a semantic slippage that arguably distorts the Nagelian elucidation of consciousness. Indeed, the phenomenal character and the phenomenality of the specific mental state of a particular subject are not, in any meaningful sense, *components* of the phenomenal character of that type of mental state. To take a specific example, *what* it is like *for Mary* to see a blue sky (Kriegel's 'qualitative character') and there being *something* it is like *for Mary* to see a blue sky (Kriegel's 'subjective character') are not *components* of *what* it is like *for any subject* to see a blue sky (Kriegel's 'phenomenal character').

In a footnote, Kriegel qualifies his analysis of the distinction between 'qualitative character' and 'subjective character' in the following terms:

There is a question as to whether the qualitative [character] and the subjective [character] are really separable. Even if they are not separable in reality, however, they are certainly separable 'in thought'. That is, there is a conceptual distinction to draw here even if no property distinction corresponds.

Kriegel (2009, p. 8, fn. 8)

The question Kriegel asks here is whether 'qualitative character' and 'subjective character' have different extensions, that is, whether one could be instantiated without the other being instantiated. Later on, he treats this as a substantive question about consciousness that could be answered positively, although he himself favours a negative answer (see Kriegel 2009, p. 53). However, this question does not make sense if 'subjective character' merely refers to the second-order property obtained by existential quantification over the first-order property of 'qualitative character'. Indeed, it is absurd to ask whether a mental state could make *some specific* contribution to its subject's overall phenomenology *without* making *any* contribution *at all* to its subject's overall phenomenology; and it is equally absurd to ask whether a mental state could make *any* contribution *at all*

to its subject's overall phenomenology *without* making *some specific* contribution to its subject's overall phenomenology.

It is worth dwelling for a moment on the term 'for-me-ness' that a number of authors have appropriated.<sup>1</sup> Zahavi and Kriegel (2016) introduce two interpretations of the term as follows:

According to a deflationary interpretation, [for-me-ness] consists simply in the experience *occurring* in someone (a 'me'). On this view, for-me-ness is a non-experiential aspect of mental life – a merely metaphysical fact, so to speak, not a phenomenological fact. The idea is that we ought to resist a no-ownership view according to which experiences can occur as free-floating unowned entities. Just as horse-riding presupposes the existence of a horse, experiencing presupposes a subject of experience. In contrast, a non-deflationary interpretation construes for-me-ness as an experiential aspect of mental life, a bona fide *phenomenal* dimension of consciousness. On this view, to say that an experience is *for me* is precisely to say something more than that it is *in me*. It is to state not only a metaphysical fact, but also a phenomenological fact. Here the relationship between experiencing and the subject goes deeper than that between horse-riding and the horse.

Zahavi and Kriegel (2016, p. 36)

Zahavi and Kriegel go on to clarify that they claim that every conscious experience has 'for-me-ness' in the 'non-deflationary' sense. Talking of experiences being 'for their subject', as opposed to merely 'in their subject', is not particularly idiomatic; however, one can see the contrast that Zahavi and Kriegel intend to mark with these phrases. The "merely metaphysical fact" associated with the deflationary interpretation of 'for-me-ness' pertains to the metaphysics of instantiation. If a vase is cracked, being cracked is a state of the vase, and the vase's instantiating the state of being cracked is a metaphysical fact about the vase. Similarly, if Paul has tacit knowledge of syntactical rules, then having knowledge of syntactical rules is a state of Paul, and Paul's instantiating of the state of having knowledge of syntactical rules is a metaphysical fact about Paul. I take it that this is what Zahavi and Kriegel mean when they appeal to the notion of a mental state being *in a subject*, which in common parlance would be equivalent to the notion of a mental state being the state *of a subject*, or alternatively to that of a subject being *in a mental state*.

<sup>1</sup>See, for example, Kriegel (2009, pp. 1-2 and p. 47), Sebastián (2012), Zahavi and Kriegel (2016), Zahavi (2014, pp. 19-24), and Zahavi (forthcoming).

Thus, the deflationary interpretation of ‘for-me-ness’ corresponds to the use of ‘for’ that specifies the the grammatical subject of an experience – namely what I labelled ‘for<sup>SUBJ</sup>’ in the context of the elucidation of the SIL and WIL phrases. ‘For<sup>SUBJ</sup> Mary to have a headache’ simply refers to the fact that Mary is the grammatical subject of the infinitive ‘having a headache’, that is, that Mary instantiates the state of having a headache. This is, indeed, a metaphysical fact about Mary. Shoemaker comments on this metaphysical fact as follows:

[A]n experiencing is something whose existence is ‘adjectival on’ a subject of experience... I am of course taking it as an obvious conceptual truth that an experiencing is necessarily an experiencing by a subject of experience, and involves that subject as intimately as a branch-bending involves a branch.

Shoemaker (1986, p. 107)

Zahavi and Kriegel’s example of horse-riding is presumably intended to make the same point: experiences are ‘adjectival’ on their subjects as horse-riding is ‘adjectival’ on horses. Thus, on the *non-deflationary* interpretation, ‘for-me-ness’ is intended to refer to a further fact about conscious mental states – a ‘phenomenological fact.’<sup>2</sup> What is the relevant phenomenological fact? One obvious answer goes as follows: in addition to the metaphysical fact that a conscious mental state *M* is a state *of* a subject *S* (or *instantiated by* a subject *S*), there is the fact that there is something it is like *for<sup>EXP</sup> S*, *for<sup>SUBJ</sup> S* to be in *M* – or equivalently, that *S*’s being in *M* constitutively contributes to *S*’s overall phenomenology. This is a ‘phenomenological’ fact in so far as it pertains to the subject’s phenomenology, although it is perhaps slightly misleading to characterise it as a *further* fact distinct from the metaphysical fact that the subject instantiates the relevant mental state. If *M* is a conscious mental state of *S*, then it is simply *in virtue* of *S*’s being in *M* that there is something it is like *for S*, *for S* to be in *M*; strictly speaking, then, it is the metaphysical fact – *S*’s *being in M*, *where M is a conscious mental state* – that contributes *ipso facto* to *S*’s phenomenology. Nonetheless, one can see how making a conceptual distinction between the relation of branch-bending to branches (or horse-riding to horses) on the one hand, and the relation of experiences to subject of experiences on the other, is warranted, helpful, and important.

<sup>2</sup>As Zahavi also puts it, “[t]he point is not simply that it is part of the very concept of experience that an experience necessarily requires an experiencer” (Zahavi 2014).

Elucidating the way in which subjects are related to their conscious mental states beyond the observation that they *instantiate* such states was precisely the purpose of the Nagelian account of consciousness offered in chapter 1.

If the non-deflationary interpretation of ‘for-me-ness’ points only to the ‘phenomenological’ fact that experiences constitutively contribute at all to their subject’s overall phenomenology, it seems innocent enough – provided that one does not suggest that ‘for-me-ness’ is a *component* of phenomenal character. However, an additional semantic shift occurs in the discussion of ‘for-me-ness’ by proponents of (CSC<sub>NE</sub>). Indeed, the notion of a mental state having ‘for-me-ness’ or being *for a subject* is often glossed in terms that suggest that the subject has a special awareness of the relevant mental state. Thus, Kriegel argues that “a mental state has subjective character just in case it is *for* the subject, in the sense that the subject has a certain awareness of it” (Kriegel 2009, p. 38). In other words, “for a conscious experience to be not only *in* me, but also *for* me, I would have to be *aware of* it” (Kriegel 2011, p. 444).<sup>3</sup> Similarly, Zahavi contends that the notion of *for-me-ness* “was introduced in order to capture the special awareness we have of our ongoing experiences” (Zahavi 2018, p. 706); indeed, “[t]he for-me-ness of experience refers to [one’s] being pre-reflectively aware of one’s own consciousness” (Zahavi 2014, p. 24).

This construal of ‘subjective character’ or ‘for-me-ness’ is echoed in a number of other publications claiming that there being something it is like for a subject to be in a mental state is a matter of the subject being conscious (or aware) *of* the mental state in some way.<sup>4</sup> Thus, the emphasis on the ‘for<sup>EXP</sup> x’ phrase in the Nagelian dictum leads these authors to the non-egological notion of self-consciousness, which in turn leads them to transition from the Nagelian dictum to (CSC<sub>NE</sub>).

<sup>3</sup>See also: “a mental state of mine is phenomenally conscious iff it has for-me-ness (subjective character); a mental state has for-me-ness (subjective character) iff I am aware of it in the right way” (Kriegel 2011, p. 444).

<sup>4</sup>The same idea has been glossed under many labels, including ‘self-consciousness’ (Kriegel 2003a; Kriegel 2004), ‘self-awareness’ (S. Gallagher 2010; Janzen 2006; Janzen 2008; Kapitan 1999; Strawson 2010; Thomasson 2006; Zahavi 1999), ‘consciousness of consciousness’ (Fasching 2008), ‘awareness of awareness’ (Montague 2016b; Strawson 2013), ‘inner awareness’ (Kriegel 2009; D. W. Smith 1986), ‘self-intimation’ (Armstrong 1968; Ryle 1949; Strawson 2013), ‘self-luminosity’ (Ganeri 2012), and ‘self-givenness’ (Zahavi 2005; Zahavi 2011). This remarkable lexical fragmentation does not help to clarify whether these authors really advocate the same claim, and if so what is the meaning of the claim in question.

As we have seen in chapter 1, Stoljar (2016) briefly discusses this transition in the form of what he calls the ‘emphatic argument’:

(P1) M is a phenomenally conscious state of S only if there is something it is like *for* S to be in M.

(P2) There is something that it is like *for* S to be in M only if S stands in some representational or awareness relation to M.

(C) M is a phenomenally conscious state of S only if S stands in some representational or awareness relation to M.

(Stoljar 2016, pp. 1193-4; with ‘X’ replaced by ‘M’)

The first premise of this argument (P1) is a consequence of the Nagelian dictum, and the conclusion (C) is a version of (CSC<sub>NE</sub>). The second premise (P2) illustrates the semantic shift that occurs through the emphasis on the ‘for<sup>EXP</sup> x’ phrase, leading to the non-egological notion of self-consciousness. Once we get to (CSC<sub>NE</sub>) through this semantic shift, one might wonder whether we have wandered away from the Nagelian elucidatory account of consciousness.

Consider the following passages taken from recent publications discussing the nature of conscious experience:<sup>5</sup>

Let us call awareness of external features and objects in one’s environment or body *outer awareness*, and awareness of internal events and states in one’s own mental life *inner awareness*... [A] mental state is phenomenally conscious only if its subject has inner awareness of it.

Kriegel (2009, p. 16)

[I]n having a visual experience of a tree in leaf (for example), the subject, in addition to being aware of the tree and any other relevant external representational content, is also aware of the awareness of the tree. In having a particular conscious perceptual experience the subject is always and necessarily also aware of that very experience itself.

Montague (2016b, p. 41)

<sup>5</sup>Most of the cited authors use the terms ‘awareness’ and ‘consciousness’ equivalently. One notable exception is Kriegel, whose notion of awareness is representational. On Kriegel’s view, to be aware of *x* is to represent *x*. Consequently, one can be *consciously* or *non-consciously* aware of something. However, unlike higher-order theorists such as Rosenthal (1990), according to which a mental state is a *conscious* mental state if and only if it is represented by a *non-conscious* higher-order mental state, Kriegel endorses the claim that the awareness we have of our conscious mental states is a *conscious* awareness, because conscious mental states *represent themselves*. Therefore, it is not inaccurate to characterize Kriegel’s view as stating that conscious mental states are states we are *conscious* of.

When I see the bottle of wine in front of me on the table, I experience (I am visually aware of) the wine bottle. But I also experience my seeing.

E. Thompson (2007, p. 285)

[W]e cannot be conscious of an object (a tasted apple, a seen chair, a touched piece of marble, a remembered event, an imagined creature, etc.) unless we are aware of the experience that allows the object to appear (the tasting, seeing, touching, remembering, imagining).

Zahavi (2014, p. 35)

[A]ll awareness involves awareness of awareness – awareness of that very awareness... To experience is of course *to experience, to be aware of*, one's experiencing... One's awareness is almost invariably focused on the world, or at least on something other than itself... but it is always also awareness of itself.

Strawson (2013, pp. 5-17)

On the face of it, all of these excerpts express a version of (CSC<sub>NE</sub>): in having a conscious experience, one is *conscious* (or *aware*) of this conscious experience. There are two main ways in which we can interpret (CSC<sub>NE</sub>). Firstly, on what we may call a *minimal* construal of (CSC<sub>NE</sub>), being *conscious* (or *aware*) of a conscious mental state M would simply mean being the subject whose overall phenomenology M constitutively contributes to.<sup>6</sup> If this is how the non-egological notion of self-consciousness is construed, then (CSC<sub>NE</sub>) does not actually go beyond the foundational claims of chapter 1.

Secondly, on what we may call a *substantive* construal of (CSC<sub>NE</sub>), being *conscious* (or *aware*) of one's conscious mental state points to some determinate aspect of phenomenology that is not already acknowledged by the foundational claims of chapter 1 as being a constitutive aspect of consciousness. As I noted, discussions of the Nagelian dictum that put a particular emphasis on the 'for<sup>EXP</sup> x' phrase appear to involve some slippage in the direction of a more substantive claim. There are two elements to this slippage: first, the suggestion that all conscious mental states have 'subjective character' or 'for-me-ness', where this notion is intended as a *component* of phenomenal character; and second, the

<sup>6</sup>Note that what I call here the 'minimal construal' of (CSC<sub>NE</sub>) involves a stronger notion than what Zahavi and Kriegel call the 'deflationary interpretation' of 'for-me-ness': what is at stake is not simply the 'metaphysical fact' that M is instantiated by S, but the 'phenomenological fact' that S's instantiation of M makes a constitutive contribution to S's overall phenomenology (as elucidated by the Nagelian account).

suggestion that ‘subjective character’ or ‘for-me-ness’ is a matter of the subject being *conscious* or *aware of* the mental state. Even if we adopt the deflationary construal of ‘consciousness of’ (or ‘awareness of’), the first element of the slippage would remain, leading us to suspect that ‘subjective character’ or ‘for-me-ness’ points to something more than what I call phenomenality.

Thus, formulations of (CSC<sub>NE</sub>) can be in principle divided into two categories. In the first category are claims that might *seem* to go beyond the foundational claims of chapter 1, but are not actually *intended* to go beyond such claims, in accordance with the *minimal* construal of (CSC<sub>NE</sub>). In the second category are claims that are actually *intended* to point to some aspect of phenomenology that is not acknowledged by the foundational claims, in accordance with the *substantive* construal of (CSC<sub>NE</sub>). Claims in the second category stand in need of further motivation as substantive claims about what is constitutive of all conscious mental states.

With this distinction in mind, I will now examine some of the most influential defences of (CSC<sub>NE</sub>) to determine whether they fall within the first or the second category. I will suggest that Zahavi’s version of (CSC<sub>NE</sub>) falls within the first category, while Kriegel’s version of (CSC<sub>NE</sub>) falls within the second category.

### 3.2 Zahavi’s minimal constitutive claim

One of the most influential contemporary accounts of the claim that consciousness constitutively involves self-consciousness in the non-egological sense (CSC<sub>NE</sub>) is due to Zahavi (1999; 2005; 2014). On Zahavi’s account, “phenomenal consciousness as such entails a thin or minimal form of self-consciousness” (Zahavi 2014, p. 14). Echoing similar remarks by Husserl and Sartre, Zahavi emphasises that this minimal form of self-consciousness is ‘pre-reflective’: it is not a matter of *reflecting* upon one’s conscious mental state. More generally, pre-reflective self-consciousness is not a matter of being in a *higher-order* mental state that represents one’s first-order conscious mental state, as claimed by higher-order theories (HOT) of consciousness (Rosenthal 1997). Rather, pre-reflective self-consciousness refers to the fact that “experience is given [to the subject], not as an object, but precisely as subjectively lived through” (*ibid.*, p. 16). This is what Zahavi means by “being pre-reflectively

aware of one's own consciousness" (*ibid.*, p. 24) and, through the emphasis on the 'for<sup>EXP</sup>  $x$ ' phrase in the Nagelian dictum, by the term 'for-me-ness':

[T]he for-me-ness... of experience simply refer[s] to the subjectivity of experience, to the fact that [one's] experiences are pre-reflectively self-conscious and thereby present in a distinctly subjective manner...

Zahavi (2014, p. 41)

How are we to understand the claim that experiences are 'present' in a 'distinctly subjective manner'? To shed light on his understanding of pre-reflective self-consciousness or 'for-me-ness', Zahavi considers the following thought experiment (Zahavi 2014, pp. 22-23; see also Zahavi 2005, p. 127):

### Phenomenal Twins

Two individuals, Mick and Mack, are physically and psychologically type-identical. Mick and Mack's respective token experiences, although numerically distinct, have exactly the same type of phenomenal character.

For the sake of clarity, consider Mick and Mack's experiences at a specific time  $t$ ; call Mick's experience at  $t$  ' $e_1$ ' and Mack's experience at  $t$  ' $e_2$ '. From a third-person perspective, Zahavi claims, there is no "relevant qualitative difference between the two [experiences]" (*ibid.*, p. 22). Having said that, Zahavi prompts the reader to abandon the third-person perspective and consider Mick's stream of consciousness from his own perspective. Consider specifically what it is like for Mick to have  $e_1$  at  $t$ . According to Zahavi, there is for Mick "a crucial difference between [ $e_1$  and  $e_2$ ], a difference that would prevent any kind of conflation [between  $e_1$  and  $e_2$ ]" (*ibid.*, p. 22). This difference consists in the fact that only  $e_1$ , and not  $e_2$ , is "given first-personally to [Mick] at all, and therefore part of [Mick's] experiential life" (*ibid.*, p. 22). To use Zahavi's alternative terminology, Mick is pre-reflectively conscious of  $e_1$ , but *not* of  $e_2$ .

One should be careful not to read too much into the claim that there is a 'crucial difference', from Mick's point of view, between his experience and Mack's. In accordance with the elucidatory account of consciousness offered in chapter 1, we can express the relevant difference as follows: while there is (constitutively) something it is like for<sup>EXP</sup> Mick, for<sup>SUBJ</sup> Mick to have  $e_1$ , there is (constitutively) nothing it is like for<sup>EXP</sup> Mick, for<sup>SUBJ</sup>

Mack to have  $e_2$ . More straightforwardly: only Mick's experience  $e_1$ , and not Mack's experience  $e_2$ , is such that it constitutively contributes to Mick's overall phenomenology. Importantly, this does not entail that pre-reflective self-consciousness or 'for-me-ness' is a special qualitative property of the experience that is somehow unique to the experience's subject, that is, that Mick's experience has a distinctive flavour of 'for-Mick-ness'.<sup>7</sup> One can see how, in this respect, Zahavi's suggestion that "for Mick, his experience will be quite unlike Mack's experience (and vice versa)" (*ibid.*, p. 24) could be misunderstood: given that Mick and Mack's experiences are type-identical (i.e., they share the exact same type of phenomenal character), what it is like for<sup>EXP</sup> Mick, for<sup>SUBJ</sup> Mick to have  $e_1$  is *exactly the same* as what it is like for<sup>EXP</sup> Mack, for<sup>SUBJ</sup> Mack to have  $e_2$ .

In the movie *Being John Malkovich* by Spike Jonze (1999), the main character is able to access John Malkovich's stream of consciousness by crawling through a mysterious portal in his office. Imagine that by using a similar science-fiction device, you could randomly access either Mick's or Mack's stream of consciousness at  $t$ . You would be completely unable to know *whose* stream of consciousness you were accessing (i.e., whether you inhabited *Mick's* stream of consciousness and lived through  $e_1$ , or whether you inhabited *Mack's* stream of consciousness and lived through  $e_2$ ). Furthermore, after repeated use of the device, you would have no way of finding out that you were successively accessing *two* numerically distinct streams of consciousness without prior knowledge of that fact. Even if you rapidly alternated between the two streams of consciousness, you would not be able to notice any change in your experience. Thus, what distinguishes  $e_1$  from  $e_2$  is merely that  $e_1$  constitutively contributes to Mick's overall phenomenology (and not Mack's), while  $e_2$  constitutively contributes to Mack's overall phenomenology (and not Mick's). There is certainly nothing in the *type* phenomenal character of  $e_1$  that distinguishes it from  $e_2$ , and vice versa.

<sup>7</sup>This echoes Parfit's remark that the only distinction between two qualitatively identical experiences is not some unique subject-specific property, but the relationship that each experience bears to its subject: "I agree with Madell that I and he could have two simultaneous experiences that were qualitatively identical, but were straightforwardly distinct. But this need not be because one of the experiences has the unique property of being mine, and the other has the unique property of being Madell's. It could simply be because one of these experiences is this experience, occurring in this particular mental life, and the other is that experience, occurring in that other particular mental life. These two mental lives might have to be referred to publicly through their connections to a pair of different human bodies." (Parfit 1984, p. 516).

I take it that Zahavi would agree with this commentary. The purpose of the **Phenomenal Twins** thought experiment, as I understand it, is to stress the fact that being the subject of a conscious mental state is not *merely* a matter of instantiating a property like any other. In particular, it differs substantially from instantiating a property like mass. Suppose that Mick and Mack, being physically type-identical, both weigh 80kg. The property of having a mass of 80kg is ‘adjectival’ on a physical object or person: there can be no having a mass of 80kg *without* something or someone whose mass it is. By contrast, pre-reflective self-consciousness or ‘for-me-ness’, in Zahavi’s sense, does not merely point to the fact that having an experience is ‘adjectival’ of a subject of experience; in other words, Zahavi refuses what Zahavi and Kriegel (2016) call the ‘deflationary interpretation’ of ‘for-me-ness’.

This is the first lesson to be drawn from the remark that the difference between  $e_1$  and  $e_2$  is *more* than a mere numerical distinction between the two experiences. Mick’s having a mass of 80kg is a different token instantiation from Mack’s having a mass of 80kg, but there is a mere numerical distinction between the two token instantiations. Mick’s having  $e_1$  is also a different token instantiation from Mack’s having  $e_2$ , but there is a sense in which the difference goes further than this ‘metaphysical’ fact. Indeed, there is something it is like for<sup>EXP</sup> Mick, for<sup>SUBJ</sup> Mick to undergo  $e_1$ , and there is something it is like for<sup>EXP</sup> Mack, for<sup>SUBJ</sup> Mack to undergo  $e_2$ . By contrast, there is nothing it is like for<sup>EXP</sup> Mick, for<sup>SUBJ</sup> Mick to have a mass of 80kg, and there is nothing it is like for<sup>EXP</sup> Mack, for<sup>SUBJ</sup> Mack to have a mass of 80kg.

The second lesson to be drawn from the thought experiment is that each conscious experience constitutively contributes to the overall phenomenology of *one and only one* subject of experience – what I have expressed in chapter 1 with the foundational claims (F5) and (F6). This is what is occasionally called (somewhat contentiously) the ‘privacy’ of experience (e.g., Sprigge 1969). Zahavi’s emphasis on this point is apparent from the following elucidation of ‘for-me-ness’:<sup>8</sup>

[T]o highlight the for-me-ness of experience is not merely to make a grammatical or logical or metaphysical point. The point is not simply that it is part of the very concept of experience that an experience necessarily requires an experiencer. No, the point

<sup>8</sup>See also Zahavi (2014, p. 22).

being made is also phenomenological. To speak of the for-me-ness of experience is to pinpoint something with ramifications for the subject's overall phenomenology. The for-me-ness of experience refers to the first-personal character of experience, to the fact that our acquaintance with our own experiential life differs from the acquaintance we have with the experiential life of others and vice versa. This difference in acquaintance or access obtains, not only when we reflect or introspect, but whenever we pre-reflectively live through an experience.

Zahavi (2014, p. 24)

Thus, on Zahavi's account, to say that I am pre-reflectively conscious of a mental state *M* – or, equivalently, to say that *M* has 'for-me-ness' – is to say (a) that *M* constitutively contributes to my overall phenomenology, and (b) that *M* does *not* constitutively contribute to any other subject's overall phenomenology. Given this minimal understanding of pre-reflective self-consciousness or 'for-me-ness', the claim that a mental state is a conscious mental state if and only its subject is pre-reflectively conscious of it – that is, if and only if it has 'for-me-ness' – strikes me as true, because it does not go beyond the foundational claims about consciousness outlined in chapter 1.

Zahavi himself emphasises that one should not understand the notion of 'pre-reflective self-consciousness' on an act-object model, similarly to the notion of 'consciousness of an object': "the experience [of an apple] is not itself an object on a par with the apple, but instead constitutes the very access to the appearing apple" (Zahavi 2014, p. 35). The relevant notion of pre-reflective self-consciousness must be understood in a minimal way, so that what it is for *S* to be (pre-reflectively) conscious of mental state *M* is for *M* to be a conscious mental state of which *S* is the subject; or, equivalently, for there to be something it is like for<sub>EXP</sub> *S*, for<sub>SUBJ</sub> *S* to be in *M*; or, also equivalent, for *M* to constitutively contribute to *S*'s overall phenomenology (to the exclusion of anyone else's). Furthermore, with these comes along *S* being in a position to know, or rationally to believe, that *S* [SELF] is in *M*.

Nonetheless, one can see how the notion of pre-reflective self-consciousness could potentially be mistaken for a less minimal notion. Using the term 'self-consciousness' and its cognates to point to notions that can be constructed within the scope of the foundational claims chapter 1 requires a considerable amount of exposition and qualification (e.g., with adjectives such as 'pre-reflective', as well as lengthy glosses). Moreover, despite Dan Zahavi's carefulness in clarifying what he intends by 'self-consciousness', his view has often

been misconstrued as pointing to a more substantive notions.<sup>9</sup> Given these two observations, one might suggest that it would be preferable – both for proponents of the minimal interpretation of (CSC<sub>NE</sub>) themselves and for their readers – to avoid the terminology of ‘self-consciousness’ altogether when discussing foundational claims about consciousness.

### 3.3 Kriegel’s substantive constitutive claim

Another very influential version of (CSC<sub>NE</sub>) is due to Kriegel (2009). Recall that Kriegel defines the ‘subjective character’ or ‘for-me-ness’ of a mental state as “what makes [it] a phenomenally conscious state *at all*” (*ibid.*, p. 1). He subsequently argues that “it is central to subjective character that it enables an epistemic or mental relation between the subject and her experience” (*ibid.*, p. 105) This epistemic or mental relation is the relation of *awareness*. To distinguish the subject’s awareness of worldly objects from the subject’s awareness of her experience itself, Kriegel calls the former *outer awareness* and the latter *inner awareness*. The specific version of (CSC<sub>NE</sub>) to be elucidated is what he calls the ‘Ubiquity of Inner Awareness Thesis’, henceforth (UIA):

(UIA) For any mental state M of a subject S, if M is conscious at *t*, then (a) S is aware of M at *t* and (b) S’s awareness of M is part of S’s overall phenomenology at *t*.<sup>10</sup>

(UIA)’s clause (b) introduces the specification that the subject’s awareness of her conscious mental state is itself part of the subject’s overall phenomenology. This idea comes through in other passages of the book:<sup>11</sup>

[I]nner awareness is such that [we] bear some epistemic relation to it, in virtue of which it is *phenomenologically manifest* to us.

Kriegel (2009, p. 182, my emphasis)

[A]nother constant element *in the fringe of consciousness* is awareness of one’s concurrent experience.

Kriegel (2009, p. 49, my emphasis)

<sup>9</sup>See Zahavi (2017; 2018) for a defence against misunderstandings of his minimal interpretation of (CSC<sub>NE</sub>).

<sup>10</sup>Kriegel (2009, p.181).

<sup>11</sup>See also: “the awareness of our conscious states is something we *experience*” (Kriegel 2003a, p. 120).

[I]nner awareness... is a *component* of the [subject's] *phenomenology*...

Kriegel (2009, p. 188, my emphasis)

Does (UIA) go beyond foundational claims about consciousness? Recall that according to the elucidation of consciousness provided in chapter 1, it is the subject's *being in a mental state* that contributes to her overall phenomenology. By contrast, (UIA) states that the subject's *being aware of the mental state that she is in* is part of her overall phenomenology. A proponent of (UIA) might say that what it is for you to be aware of a mental state simply is for you to *be* the experiencing subject of that mental state. Accordingly, (UIA) would mean that for any mental state M of a grammatical subject S, if M is a *conscious* mental state, then (a) S is the experiencing subject of M, and (b) S's being the experiencing subject of M is part of S's overall phenomenology. Suppose that you are (perceptually) aware of a yellow lemon. On this interpretation of (UIA), you have an (inner) awareness of your (outer) awareness of the yellow lemon; but your being aware of your awareness of the yellow lemon would be nothing more than your being the experiencing subject of your awareness of the yellow lemon. This interpretation does not seem to point to aspects of phenomenology that are not already acknowledged by foundational claims about consciousness as being constitutive of consciousness.

However, this is not how Kriegel himself seems to interpret (UIA). As we have seen, Kriegel suggests that the 'subjective character' or 'for-me-ness' of all conscious mental states is a *component* of their phenomenal character (Kriegel 2009, p. 116). Since Kriegel claims that the 'subjective character' or 'for-me-ness' of a conscious mental state is equivalent to the subject's *inner awareness* of that mental state, this suggests that such inner awareness is somehow a *component* of the state's phenomenal character; on this view, your awareness of your awareness of a yellow lemon would be a part of what it is like for you to be aware of the yellow lemon.

In a more recent publication, Kriegel makes it clear that his understanding of (UIA) is intended to go beyond any of the foundational claims discussed in chapter 1. Indeed, Kriegel stresses that the relation between inner awareness or 'for-me-ness' and phenomenal character is not exhausted by the point that the former is a determinable of which the latter is a determinate:

While the bluish way it is like for me to see blue is different from the reddish way it is like for me to see red, the element of *for-me-ness* in these two ways-it-is-like-for-me is strictly identical, and not only in the sense that we can define a genus, or determinable, of which both bluishness and reddishness are species, or determinates, and which *qua* genus or determinable remains invariant. Rather, there is a very specific, very determinate aspect of bluish-for-me-ness and reddish-for-me-ness that is common to the two, namely, *for-me-ness* as such.

Kriegel ([forthcoming](#))

Kriegel is clear that inner awareness or ‘for-me-ness’ is not merely a ‘pre-condition’ for a mental state to contribute to the subject’s phenomenology, but it is also a *distinct* component of the subject’s phenomenology. Thus, Kriegel compares the relation between inner awareness and the subject’s phenomenology to the relation between a keystone and thirteen-stone masonry arch:

As a mere commonality and yet a substantive one, *for-me-ness* serves a double function as both (i) a component among others in a conscious state’s overall phenomenal character and (ii) a precondition for the existence of all other phenomenal components (*as* phenomenal components). Compare the keystone of a thirteen-stone masonry arch. On the one hand, it is a stone among others composing the arch, as intrinsically ‘beefy’ as the other twelve. On the other hand, if we remove it the whole arch collapses, and to that extent it is a precondition for there being any other arch-component.

Kriegel ([forthcoming](#))

By analogy with the keystone, your (inner) awareness of your (outer) awareness of a yellow lemon is taken to both (a) enable your (outer) awareness of the yellow lemon to make any contribution *at all* to your phenomenology, and (b) make its own, *independent* contribution to your phenomenology. The resulting version of the constitutive claim is clearly substantive, and does not plausibly follow from the Nagelian elucidatory account of consciousness. It stands in urgent need of support.

Kriegel himself suggests that “the best reason to think that there is such as a thing as... inner awareness is phenomenological: [it] is simply phenomenologically manifest” (2009, p. 50). Thus, he takes his view to be phenomenologically plausible, so much so that he “cannot envisage what it would be like to have a phenomenology lacking the kind of inner awareness that constitutes *for-me-ness*” (2009, p. 175). However, many authors

have attested that they do not share this intuition.<sup>12</sup> If inner awareness were indeed a ubiquitous and phenomenologically manifest feature of consciousness, one might expect a broader agreement on its existence. Kriegel readily admits that “[t]he awareness in question is certainly somewhat elusive” (2009, p. 47). He elaborates on the elusiveness of inner awareness or ‘for-me-ness’ through two additional claims that we can formulate as follows:

(FMN1) Most of the time, inner awareness or ‘for-me-ness’ is *peripheral*, rather than *focal*.

(FMN2) Peripheral inner awareness can never be *introspected*.

Let us unpack each of these claims, starting with (FMN1). Kriegel introduces the distinction between *focal* and *peripheral* awareness by analogy with the structure of the visual field. Human vision has a centre/periphery structure in so far as visual stimuli are perceived with greater clarity in the fovea than in the peripheral visual field. By analogy, Kriegel argues that experience itself has a centre/periphery structure: some aspects of one’s overall phenomenology are *focal*, while others are *peripheral*. This analogy is not to be taken literally, because the centre/periphery structure of experience is obviously not a matter of physiology (e.g., the density and sensitivity of receptor cells such as those covering the retina); rather, it is meant to be determined by the allocation of attentional resources.<sup>13</sup> Indeed, Kriegel glosses the distinction between *focal* and *peripheral* awareness as a distinction between *attentive* and *inattentive* awareness.<sup>14</sup> Thus, *focal* awareness is the awareness of what we attend to, while *peripheral* awareness is the awareness of what we do not attend to.

According to Kriegel, the distinction between *focal* and *peripheral* awareness cuts across the distinction between *outer* and *inner* awareness. Outer awareness of external objects can be either focal (e.g., when one attends to the yellow lemon one sees) or peripheral (e.g., when one sees a yellow lemon without attending to it), and most experiences involve both varieties of outer awareness at the same time (e.g., one may be *focally* aware of the yellow

<sup>12</sup>See, for example, Dretske (1993), Gennaro (2008), Gertler (2012), Howell and B. Thompson (2017), McClelland (2015), Mehta (2013), Siewert (1998), and D. W. Smith (2004).

<sup>13</sup>See Kriegel (2009, p. 183).

<sup>14</sup>“The distinction that is relevant to our present purposes is between attentive/inattentive... awareness”, Kriegel (2009, p. 183, n. 20).

lemon while being *peripherally* aware of the hum of the refrigerator in the background). Inner awareness, in turn, is understood to be typically peripheral: in having an experience, one is aware of it even if one does not attend to it. This is the first part of Kriegel's answer to sceptics – if you doubt that your experience constantly involves a phenomenologically manifest inner awareness or 'for-me-ness', this might be because such inner awareness or 'for-me-ness' is typically peripheral.

Kriegel goes on to argue that peripheral inner awareness *becomes* focal when one introspects (attends to) one's occurrent experience. Thus, "introspective awareness is the same old constitutive, inbuilt awareness that... every conscious state involves, though properly augmented with additional attention resources" (Giustina and Kriegel 2017, p. 158). This is what Kriegel calls the 'attention-shift' model of introspection: introspection consists in shifting attention from external objects to one's awareness of these objects, through a conversion of *peripheral* inner awareness into *focal* inner awareness. Introspection, on this view, is nothing but an inbuilt inner awareness of one's experience to which attentional resources have been allocated.

This model of introspection leads to the second part of Kriegel's answer to sceptics: since introspection results from the conversion of peripheral inner awareness into focal inner awareness, it follows that peripheral inner awareness itself cannot be introspected (FMN2). Indeed, when we introspect, we no longer have a *peripheral* inner awareness of our experience, and our *focal* inner awareness of our experience is the very relation of introspection. The reason why critics deny the ubiquity of inner awareness is that they fail to notice such awareness upon introspecting their experience – but this is precisely what the attention-shift model of introspection predicts: "introspecting cannot reveal peripheral inner awareness because it *annihilates* it (by supplanting it)" (Kriegel 2009, p. 184).

Kriegel argues that the 'attention-shift' model of introspection explains how (UIA) can be compatible with the so-called 'transparency of experience' thesis. The traditional formulation of the transparency thesis states that when one attempts to attend to one's experience, one can only attend to features of the worldly objects presented by one's experience (Harman 1990). In other words, attempting to introspect one's experience does not reveal properties of the experience itself, but only properties of its objects – properties of what

it is an experience *of*. Note that this formulation of the transparency thesis is not strictly incompatible with (UIA), because (UIA) does not explicitly say that inner awareness is ever *focal* rather than *peripheral*. However, Kriegel does need to give some account of what happens when we attend to our occurrent experience, that is, when our *peripheral* inner awareness of that experience is supplanted by the *focal* inner awareness of it.

Kriegel denies that his view is incompatible with the transparency thesis, which he formulates in the following way:

(TE) For any experience  $e$  and any feature  $F$ , if  $F$  is an introspectible feature of  $e$ , then  $F$  is (part of) the first-order representational content of  $e$ .<sup>15</sup>

The notion of first-order representational content refers here to content that does not involve representations or representational properties, but only involves worldly objects and their properties. In other words, (TE) states that the only features of one's experience that are available to introspection are those that are part of the experience's world-directed representational content. Kriegel goes on to say that his view is compatible with (TE), because the attention-shift model predicts that *peripheral* inner awareness cannot be introspected. Upon introspecting, peripheral inner awareness turns into focal inner awareness, and such awareness is directed to the first-order representational content of the experience.

This treatment of the transparency of experience raises several concerns. Firstly, (TE) does not reflect traditional formulations of the transparency thesis. Indeed, the very formulation of (TE) suggests that one can introspect properties of one's experience, which is precisely what the transparency thesis is normally taken to deny. Moreover, it is difficult to see how a property could be both the property of an experience, and part of what Kriegel calls the experience's 'first-order representational content'. Suppose that you have a visual experience of the blue sky. Kriegel might say that the blueness of the sky is part of the 'first-order representational content' of that experience, insofar as the experience represents the sky as blue; however, the blueness of the sky is a property of the sky, not a property of your experience of it. The corresponding property of the experience would be the property of

<sup>15</sup>Kriegel (2009, p. 181); see also Kriegel ([forthcoming](#)).

representing (or presenting) blueness – but such a property is not part of the ‘first-order representational content’ of the experience in Kriegel’s sense. According to the traditional version of the transparency thesis, any attempt to attend to your experience of the blue sky would result in attending to what your experience represents, namely the blue sky itself. Interestingly, Kriegel himself initially introduces the transparency thesis as the claim that “whenever we try to introspect one of our experiences, we can become acquainted only with what it is an experience of – not with the experiencing itself” (Kriegel 2009, p. 69). This is presumably not equivalent to (TE).

A more straightforward and accurate formulation of the transparency thesis would be the following:

(TE\*) For any experience  $e$  (e.g., a visual perceptual experience), upon attempting to attend to a feature (i.e., a property) of  $e$ , one can only attend to objects and properties that are part of the worldly scene that  $e$  represents.

The formulation of the transparency thesis as (TE\*) suggests that it is in fact *incompatible* with Kriegel’s account. Consider again the experience of looking at the blue sky. According to Kriegel, “although I am not focally aware of the experience itself, but rather of the sky, I am nonetheless peripherally aware of the experience itself. That is, the experience combines focal outer awareness of the sky with peripheral inner awareness of itself.” (Kriegel 2009, p. 17). Now, suppose that you introspect your experience of the blue sky. According to the attention-switch model, this consists in reallocating your attentional resources, such that your experience of the blue sky becomes the focus of your attention (fig. 3.1). The same inner awareness that was before *peripheral* is now *focal*, and introspection simply consists in having such focal inner awareness of one’s experience. If introspection is just focal inner awareness, and focal inner awareness is just peripheral inner awareness plus attention, then it would seem that anything of which a subject has *peripheral* inner awareness is available to introspection (although not as peripheral) – given that it is a possible target for *focal* inner awareness.

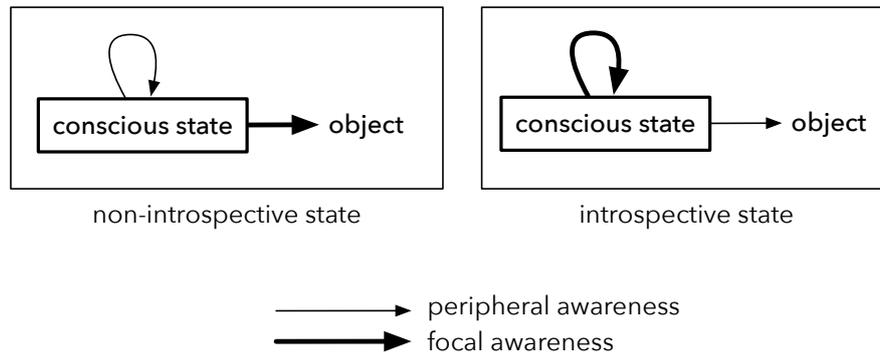


Figure 3.1: The attention-shift model of introspection.

It follows that there must be something that is available to introspection over and beyond the experience's first-order representational content. If not, then (UIA) would become vacuous, for there would be no difference between outer awareness and inner awareness. Because the focal inner awareness that constitutes introspection just is the same in-built inner awareness that happens to be peripheral in the non-introspective case, introspecting one's experience cannot be the same thing as attending to the object and properties it represents. To see this, consider the two awareness relations in Kriegel's model: inner awareness (the loopy arrow in fig. 3.1) or outer awareness (the straight arrow in fig. 3.1). We know that each of these awareness relations can be qualified by the allocation of attentional resources, to be either focal (attentive) or peripheral (inattentive). Furthermore, we know that the introspection relation can only be of the focal (attentive) variety. Upon introspecting, one of the two awareness relations – the *inner* awareness relation – becomes *focal*, and this is what introspection consists in (see Kriegel 2009, p. 184).

If, upon introspecting one's experience, it was only possible to attend to objects and properties that are part of the worldly scene that the experience represents – as stated by (TE\*) –, it would suggest that the introspection relation is in fact a focal *outer* awareness relation. In turn, this would mean that peripheral *inner* awareness no longer has a role as a 'precursor' to introspection in non-introspective states, and the real 'precursor' to introspection would be peripheral *outer* awareness (inattentive awareness of external objects). Such a view would leave the nature of peripheral inner awareness wholly mysterious.

Thus, Kriegel's account combining (UIA) and the 'attention-shift' model of introspection does not seem compatible with the transparency thesis, if the latter is adequately formulated as in (TE\*). Given that Kriegel takes the transparency thesis to be phenomenologically plausible (e.g., Kriegel 2009, pp. 69-71), this conflict undermines the weight of his phenomenological intuition in favour of (UIA).

Admittedly, Kriegel does not pretend to offer a positive argument in favour of his interpretation of (UIA); rather, he states his view of inner awareness as an "unpedestrian phenomenological pronouncement" (*ibid.*, p. 166), and subsequently defends it against some potential objections, and in particular the objection that it is incompatible with the transparency thesis. This argumentative sequence is a part of Kriegel's broader project. Indeed, Kriegel argues that the only theory of consciousness that can accommodate his interpretation of (UIA) is his preferred theory, called self-representationalism, according to which what makes a mental state M of S a conscious mental state is that it *represents itself* – thereby making S aware of M.<sup>16</sup> In the context of this broader project, Kriegel admits that his interpretation of (UIA) is both substantive and controversial:

I consider it *a substantive claim*, not a matter of definition, that subjective character amounts normally to peripheral inner awareness.

Kriegel (2009, p. 47, fn. 38, my emphasis)

My claim, and *I do not mean it as trivial or uncontroversial*, is that subjective character just is normally peripheral inner awareness.

Kriegel (2009, p. 50, my emphasis)

Nonetheless, Kriegel claims not only that his view is phenomenologically plausible, but also that his rebuttal of potential objections should lead us to conclude that "we have *no good reason* not to accept that the overall phenomenology of a conscious subject at a time always and necessarily includes an element of inner awareness" (*ibid.*, p. 196).

I have argued that we may in fact have good reasons not to endorse Kriegel's interpretation of (UIA). Beyond the anecdotal observation that I – like many others – find

<sup>16</sup>More precisely, on the self-representationalist theory, any conscious mental state M has two components, M<sub>1</sub> and M<sub>2</sub>, such that M<sub>2</sub> directly represents M<sub>1</sub>, and thereby indirectly represents M itself. The metaphysical details of this theory are complex and need not detain us here (see McClelland [forthcoming](#) for an overview).

Kriegel's account particularly implausible as a characterisation of what is 'phenomenologically manifest' in every conscious experience, I suggested that it is incompatible with the transparency thesis that Kriegel himself seemingly takes for granted. More importantly, the transition from the Nagelian dictum to Kriegel's inflationary interpretation of (UIA), through the emphasis on the 'for me' phrase, seems to involve a form of double-counting that equivocates on the meaning of 'subjective character' or 'for-me-ness'. Consider the challenge that Kriegel addresses to critics of his view:

[P]hilosophers [who deny] the very existence of for-me-ness... owe us an alternative account of the substantive commonality among conscious states, or an argument to the effect that there is no substantive commonality among conscious states. In that respect, for-me-ness is not just phenomenologically compelling, but also does a certain explanatory work, insofar as it accounts for the apparent substantive commonality across experiences.

Kriegel ([forthcoming](#))

On the view I defended following Nagel's elucidation of consciousness, the 'substantive' commonality among conscious mental states is merely the second-order property of phenomenality: all conscious mental states are constitutively such that they make *some* contribution *or other* to their subject's overall phenomenology.<sup>17</sup> The second-order property of phenomenality is certainly not a *component* of the first-order property of phenomenal character in any meaningful sense. In other words, what makes a mental state M of a subject S a *conscious* mental state *at all* is not *part of* what it is like for S to be in M. To treat phenomenality as a component of phenomenal character is not only phenomenologically implausible – in so far as it is implausible that “for-me-ness [is] ubiquitous in conscious experience [and is] *necessarily so*” (Kriegel [forthcoming](#)) –, but perhaps also metaphysically suspect – in so far as a second-order property obtained by existential quantification over a first-order property cannot be a component of the first-order property.

In summary, Kriegel's version of (CSC<sub>NE</sub>) explicitly goes beyond the foundational claims laid out in chapter 1: it is a substantive and controversial claim that does not plausibly follow from the Nagelian elucidatory account of consciousness. On Kriegel's account,

<sup>17</sup>Furthermore, the commonality among the conscious mental states *of a particular subject S* is that they all make some contribution or other to S's overall phenomenology.

being in a conscious mental state constitutively involves having a (typically peripheral) inner awareness of that state. Such inner awareness is phenomenologically manifest; although we cannot introspect it, “we do have a *general impression* of peripheral inner awareness from our ordinary, non-introspective consciousness” (Kriegel 2009, p. 186). However, Kriegel does not give a positive account of what this ‘general impression’ of our inner awareness consists of.<sup>18</sup> Ultimately, his version of (CSC<sub>NE</sub>) stands in need of further support as a substantive claim about what is constitutive of all conscious mental states. Additionally, and perhaps more worryingly, his account is potentially misleading in its treatment of ‘subjective character’ or ‘for-me-ness’ as a *component* of phenomenal character.

### 3.4 Conclusion

In this chapter, I examined the first variant of (CSC), according to which consciousness constitutively involves self-consciousness in the non-egological sense (consciousness of *one’s experience*). On the minimal interpretation of (CSC<sub>NE</sub>) adopted by Zahavi, the non-egological notion of self-consciousness points to aspects of phenomenology also acknowledged by the foundational claims of chapter 1. Once it has been properly understood, this minimal interpretation of the claim is both plausible and illuminating; but it is preferable to formulate foundational claims about consciousness without using a term as polysemous as ‘self-consciousness’, which is likely to invite a stronger reading. On the more substantive interpretation of (CSC<sub>NE</sub>) favoured by Kriegel, the non-egological notion of self-consciousness (or ‘inner awareness’) is treated both as the second-order property of phenomenality and as a first-order component the phenomenal character of mental states. The resulting claim stands in need of further support, and is potentially misleading. The upshot of this discussion is that we should avoid treating the non-egological notion of self-consciousness as a constitutive aspect of the notion of all conscious mental states.

<sup>18</sup>See, for example, Kriegel (2009, p. 186).

## 4

# Debunking the Myth: Egological Self-Consciousness

Self-consciousness appears, in many cases, as an intermittent mode of conscious experience.

Titchener (1911, pp. 551-2)

In the previous chapter, I have argued that (CSC<sub>NE</sub>), the claim that consciousness constitutively involves consciousness *of one's experience*, stands in need of support if it is intended as a substantive claim about determinate aspects of phenomenology. Given the Nagelian account of consciousness, it is not conceptually possible for a subject to have a conscious experience that does not constitutively contribute to the subject's overall phenomenology. However, it is at least conceptually possible for a subject to have a conscious experience without being conscious *of that experience*, in so far as this notion goes beyond the scope of the Nagelian elucidatory account of consciousness and of the foundational claims of chapter 1.

In this chapter, I will turn to the second constitutive claim (CSC<sub>E</sub>), according to which consciousness constitutively involves consciousness *of oneself*:

(CSC<sub>E</sub>) Necessarily, for any subject S, S is *conscious at t* if and only if S is conscious *of S* at *t*.

My overall strategy will follow that of the previous chapter. In §4.1, I examine how some proponents of (CSC<sub>E</sub>) gloss the notions of 'subjectivity', 'subjective character' and 'form-ness' to arrive at the claim that consciousness constitutively involves consciousness *of oneself*. I subsequently argue, once again, that formulations of (CSC<sub>E</sub>) may be divided

into two categories. In the first category are claims that point to aspects of phenomenology that are acknowledged by foundational claims about consciousness. I examine the claims that fall within this first category in §4.2. In the second category are claims that point to *determinate* aspects of phenomenology that are not acknowledged by the foundational claims as being constitutive aspects of consciousness. I examine the claims that might fall within this second category in §4.3, and argue that they lack argumentative support. This is where the parallel between this chapter and the previous one ends. In §4.4, I consider a weaker but still substantive claim about egological self-consciousness, according to which being conscious *typically* or *ordinarily* involves a form of consciousness of *oneself*. I suggest that while this claim is antecedently plausible, a full assessment would involve distinguishing between more determinate notions of egological self-consciousness and examining empirical evidence regarding their prevalence in conscious experience. This will be the purpose of the second part of this thesis.

#### 4.1 From subjectivity to consciousness of *oneself*

In the last chapter, we have seen how glosses of the Nagelian dictum could lead some authors, through an emphasis on the ‘for *x*’ phrase, to the non-egological notion of self-consciousness (consciousness of *one’s experience*). Interestingly, some authors use the same emphasis to arrive at what seems like an *egological* notion of self-consciousness (consciousness of *oneself*). Consider, for example, the following passage:<sup>1</sup>

Pre-reflective self-consciousness specifically corresponds to consciousness of the self as it is the *subject* of any given experience. Compare two different experiences: the smelling of fresh coffee and the seeing of midnight sun. These experiences differ in their phenomenality, i.e. in ‘what it feels like’ to undergo them... However, these experiences... share a specific dimension in the fact that they are all given from the first-person perspective, they are given (at least tacitly) as *my* experiences, as experiences *I* am undergoing: they feel like something *for me*. This quality of *mineness* or *for-me-ness* is what the notion of pre-reflective consciousness of the self-as-subject designates.

Legrand (2007, p. 584)

<sup>1</sup>See also Sebastián (2012, pp. 160-2) and Duncan (2018, p. 88), among other examples.

This passage is reminiscent of the exposition of ‘subjectivity’, ‘subjective character’ and ‘form-ness’ discussed in the previous chapter. However, it marks a different emphasis on the subject of an experience being conscious *of herself, as the subject* of that experience.

Being conscious *of oneself as the subject* of one’s occurrent experience is *prima facie* not equivalent to being conscious *of that experience*. The former might entail the latter, at least on the following interpretation: if a subject S in conscious mental state M is conscious of S as the subject of M, then (a) S is conscious of the relation that holds between S and M, and (b) S is conscious of the *relata* of that relation, namely (i) S and (ii) M. However, being conscious *of one’s experience* presumably does not conversely entail being conscious *of oneself as the subject* of that experience.

Both Zahavi and Kriegel acknowledge the difference between the non-egological notion of self-consciousness at play in (CSC<sub>NE</sub>) and the egological notion of self-consciousness at play in (CSC<sub>E</sub>). Thus, Zahavi writes that “[b]eing pre-reflectively aware of one’s experiences is neither tantamount to being aware of oneself as an object, nor equivalent to being thematically aware of the experiences *as one’s own*” (Zahavi [forthcoming](#)). Kriegel himself admits being tempted by a claim close to (CSC<sub>E</sub>), but falls short of endorsing the idea that consciousness *of oneself* is constitutive of consciousness:<sup>2</sup>

I would say that my current experience’s pre-reflective self-consciousness strikes me as egological – that is, as a form of peripheral *self*-awareness. My peripheral awareness of my current experience is awareness of it *as mine*. There is an elusive sense of self-presence or self-manifestation inherent in even a simple conscious experience of the blue sky. It is less clear to me, however, that this feature of peripheral inner awareness – its being *self*-awareness and not *mere* inner awareness – is *constitutive* of phenomenology.

Kriegel (2009, p. 177)

As we have seen, however, some formulations of (CSC<sub>NE</sub>) point to aspects of phenomenology already acknowledged by the Nagelian elucidatory account of consciousness and the foundational claims that plausibly follow from it. Thus, Zahavi’s account of pre-reflective self-consciousness ultimately boils down to the claim that conscious mental states are constitutively such that they contribute to their own subject’s overall phenomenology,

<sup>2</sup>Kriegel does seem to endorse (CSC<sub>E</sub>) in his earlier (2003b, pp. 13-21).

and to no other subject's overall phenomenology. The above passage from Legrand (2007) also presents itself as an account of 'pre-reflective self-consciousness'; consequently, one might wonder whether her version of (CSC<sub>E</sub>) also boils down to one or several of the foundational claims considered in chapter 1. Despite appearances to the contrary, there might not be any daylight between some versions of the claim that consciousness constitutively involves consciousness *of one's experience* (CSC<sub>NE</sub>), and some versions of the claim that consciousness constitutively involves consciousness *of oneself* (CSC<sub>E</sub>).

One noteworthy feature of the above passage from Legrand (2007) is that it concerns a consciousness of oneself *as the subject* of one's experience, while Zahavi (forthcoming) claims that "[b]eing pre-reflectively aware of one's experiences is [not] tantamount to being aware of oneself *as an object*" (my emphasis).<sup>3</sup> The distinction between being conscious of oneself *as a subject* and being conscious of oneself *as an object* complicates the distinction between non-egological and egological self-consciousness. As we shall see, it is not always clear that the version of (CSC<sub>E</sub>) defended by authors who put an emphasis of the subject being conscious of herself *as a subject* really differs from Zahavi's minimal interpretation of (CSC<sub>NE</sub>), and, consequently, from the foundational claims discussed in chapter 1.

With this in mind, let us consider additional examples of the claim that consciousness constitutively involves consciousness *of oneself*:

If 'self-consciousness' is taken to mean 'consciousness with a sense of self', then all human consciousness is necessarily covered by the term – there is just no other kind of consciousness as far as I can see.

Damasio (1999, p. 19)

[A]ll consciousness involves consciousness of self... [A]ll conscious states make me, at least prereflectively, aware of myself... One is always, at least prereflectively, in the presence of oneself in being present to the world.

Wider (2006, pp. 63-78)

I can recognize... a sense in which a special form of self-consciousness *is* built into the character of experience. [For example] some sort of peripheral consciousness of oneself as a situated, active perceiver... is essentially involved in the ordinary experience of looking.

Siewert (2013, p. 256)

<sup>3</sup>See also Zahavi (2018, p. 706).

[O]ne cannot consciously experience anything without thereby being aware of oneself... In having an experience we are necessarily aware... of ourselves as the one to whom something is phenomenally given.

Nida-Rümelin (2014, p. 269)

On the face of it, all of these excerpts express a version of (CSC<sub>E</sub>): in having a conscious experience, one is *conscious (or aware) of oneself*. As in the case of (CSC<sub>NE</sub>), we can divide interpretations of (CSC<sub>E</sub>) into two categories. First, on a *minimal* interpretation of (CSC<sub>E</sub>), being conscious (or aware) *of oneself* as the subject of a conscious mental state M would mean being the subject whose overall phenomenology M constitutively contributes to. If this is how (CSC<sub>E</sub>) is interpreted, then it does not actually go beyond the Nagelian elucidatory account of consciousness provided in chapter 1; nonetheless, in so far as it plausibly follows from such an account, this minimal interpretation of (CSC<sub>E</sub>) can be illuminating.

Second, on a more *substantive* construal of (CSC<sub>E</sub>), being *conscious (or aware) of oneself* whenever one is in a conscious mental state points to some determinate aspect of phenomenology that is not already acknowledged by the foundational claims of chapter 1. Specifically, it might point to some kind of *sense of self* that is part of what it is like for a subject to be conscious at all. This is arguably the most natural understanding of the expression ‘consciousness *of oneself*’.

Following the same strategy as that of the previous chapter, I will first examine versions of (CSC<sub>NE</sub>) that seem to fall within the first category, then examine versions of (CSC<sub>E</sub>) that seem to fall within the second category.

## 4.2 Minimal constitutive claims

It is not immediately clear how (CSC<sub>E</sub>) could be a claim that merely points to aspects of phenomenology already acknowledged by foundational claims about consciousness. For example, (CSC<sub>E</sub>) clearly does not follow from foundational claim (F5):

(F5) If M is a conscious mental state, then there is a subject S such that S is in M (and, consequently, S’s being in M constitutively contributes to S’s overall phenomenology).

If M is a visual experience of a yellow lemon, then there is a subject S such that S has this visual experience, and furthermore this visual experience constitutively contributes to S's overall phenomenology: there is something it is like for S to be conscious of a yellow lemon. Presumably, it does not follow that S is conscious of *herself*, or conscious of *being the subject to whom a yellow banana is presented*. It certainly seems conceivable that one could have a conscious experience that represents the world as being thus-and-so without representing oneself as the subject of that experience.<sup>4</sup>

(CSC<sub>E</sub>) does not follow either from foundational claim (F3):

(F3) If M is a conscious mental state of subject S, then S's being in M gives S a reason to judge rationally that S [SELF] is in M.

A subject S being the subject of a conscious mental state M does put S in a position to judge rationally that her being in M is an aspect of her overall phenomenology.<sup>5</sup> Furthermore, the judgement that S is thus in a position to make is not simply that *someone* is in M, but that *herself* is in M. In other words, by being in M, S is in a position to judge "I am in M". Such a judgement involves a representation of oneself *as oneself*. Consequently, if S does come to make such a judgement consciously, she will *ipso facto* be self-conscious in so far as making a conscious judgement constitutively contributes to one's overall phenomenology.<sup>6</sup>

However, being *in a position to judge* that *p* is not equivalent to judging that *p*. In normal circumstances, one does not constantly self-ascribe the conscious mental states that one undergoes. Therefore, having a visual experience of a yellow lemon does not typically involve representing oneself as the subject of that experience – that is, unless one does come to judge "I am having a visual experience of a yellow lemon".

Let us examine some of the defences of (CSC<sub>E</sub>) quoted above, starting with Legrand (2006). Initially, Legrand appears to suggest that she intends her claim to go beyond the

<sup>4</sup>As Bayne puts it, "it seems at least conceptually possible to have a phenomenal state without experiencing oneself as the subject of that phenomenal state" (Bayne 2004, p. 231).

<sup>5</sup>Note that S will not always be in a position to *know* that she is in M, because of the kind of scenario emphasised by Williamson's anti-luminosity argument; but even then, as foundational claim (F4) states, she would presumably be in a position to know that she is in M *in most cases*. See chapter 1 for a discussion.

<sup>6</sup>I will come back to the kind of self-consciousness afforded by conscious propositional attitudes in chapter 5.

Nagelian elucidation of consciousness. For example, she writes that “[i]t is important to note that the specificity of pre-reflective consciousness of the self-as-subject is not fully captured by the notion of phenomenal consciousness” (2006, p. 584). This statement is potentially misleading, however, for Legrand actually means that all conscious experiences involve pre-reflective self-consciousness even though their phenomenal character varies. Furthermore, she suggests that being pre-reflectively self-conscious is not a matter of having an experience of one’s self:

[I]t is crucial to understand that the notion of pre-reflective self-consciousness does not suppose that the self would be experienced as standing opposed to the stream of consciousness. Rather, at the pre-reflective level, it is an integral part of conscious experience... [A]ny experience is pre-reflectively experienced as intrinsically subjective in the sense that it is experienced from the perspective of the experiencing subject... The latter is a first-person perspective; it is tied to a self in the sense of being tied to the point of view of the experiencing, perceiving, acting subject.

Legrand (2007, p. 584)

This passage makes it rather clear that the relevant notion of pre-reflective self-consciousness is very minimal, and might in fact not differ at all from Zahavi’s notion of pre-reflective self-consciousness.<sup>7</sup> On this minimal understanding, the appeal to self-consciousness is meant to emphasise that the subject of an experience is not merely the *grammatical* or *instantiating* subject on which the experience is ‘adjectival’ (as *branching* is to a *branche*), but the *experiencing* subject whose overall phenomenology the experience constitutively contributes to.

Let us now turn to Nida-Rümelin’s version of (CSC<sub>E</sub>) (Nida-Rümelin [forthcoming](#); 2014; 2017; 2018). Nida-Rümelin states her view as follows:

Being phenomenally conscious (which is nothing else than instantiating experiential properties) involves being aware of having those experiential properties... which in turn involves... being aware of oneself... One might say that phenomenal consciousness and pre-reflective self-awareness are just two sides of one and the same ‘phenomenon’.

Nida-Rümelin (2018, p. 3380)

<sup>7</sup>The fact that Legrand repeatedly cites Zahavi in the passage that elucidates her notion of pre-reflective self-consciousness is further evidence of this equivalence.

Some technical exposition is needed to fully understand Nida-Rümelin's account. On her account, properties of *subjects of experience* are explanatorily prior to properties of *experiences* in providing an elucidation of consciousness. Instead of saying that *experiences* instantiate phenomenal properties, Nida-Rümelin finds it more illuminating to say that *subjects* instantiate 'experiential properties', where these are a special class of properties that are constitutively such that there is something it is like for a subject to instantiate them. As she also puts it, experiential properties are properties whose instantiation by a subject S at *t* *partially constitutes* what it is like for S at *t*. In my own preferred terminology, one might say that experiential properties are properties that constitutively contribute to the overall phenomenology of the subject who instantiates them. In turn, experiences are events that consist in the instantiation of experiential properties by a subject.

The second important element of Nida-Rümelin's account is actually formulated as a version of (CSC<sub>NE</sub>): consciousness constitutively involves "awareness of one's own experience such that it is impossible to undergo an experience without being... aware of undergoing it" (2018, p. 3368). This is what Nida-Rümelin calls 'primitive awareness'. On this account, if you have an experience of a yellow lemon, then you must have a 'primitive awareness' of this experience the yellow lemon, and such awareness is *constitutive* of your having that experience. Importantly, your primitive awareness of your experience of the yellow lemon does not seem to be anything over and beyond the fact that your experience constitutively contributes to your overall phenomenology. Nida-Rümelin also characterises primitive awareness in terms of the subject's relation to experiential properties:

Whenever a subject has an experiential property E, the subject is aware of having property E in a way which is constitutive of having E. For example, when you are phenomenally presented with a specific pain in your head, then you must, in a sense, *be aware of your being presented with that pain* in order for it to be true that the pain is phenomenally presented to you.

Nida-Rümelin (2014, p. 264)

In consciously seeing a yellow lemon, one might say that you are phenomenally presented with yellow; this involves having a 'primitive awareness' of being phenomenally presented with yellow, where such 'primitive awareness' is *constitutive* of your being phenomenally

presented with yellow. This account of primitive awareness, like Zahavi's account of pre-reflective self-consciousness, does not seem to go beyond foundational claims that follow from the Nagelian elucidation of consciousness. Indeed, Nida-Rümelin ultimately elucidates the notion of primitive awareness as follows:<sup>8</sup>

A subject is necessarily aware of being presented with something in the sense that being so presented with something makes a difference for the subject at that moment.... A subject is primitively aware of having [an experiential property] E just in case that having E makes a difference for the subject: *Having E partially constitutes the subject's overall phenomenology.*

Nida-Rümelin (2014, p. 266)

We are back in familiar territory: a property E of a subject S (e.g., being phenomenally presented with yellow) is an *experiential* property of S just in case S's instantiation of E constitutively contributes to S's overall phenomenology; and for S to be 'primitively aware' of E in Nida-Rümelin's sense is for E to constitutively contribute to S's overall phenomenology. Thus, there is a use of 'awareness of' (namely, 'primitive awareness of') that can be constructed within the scope of foundational claims about consciousness.

We can now ask how awareness *of oneself* comes into this picture to yield a version of (CSC<sub>E</sub>); and this is where it becomes more difficult to interpret Nida-Rümelin's account. In a number of passages, Nida-Rümelin suggests that primitive awareness involves a form of awareness *of oneself*, in the following sense: being aware of an experiential property is, *ipso facto*, being aware of being the subject who instantiates this experiential property – that is, being aware of being the subject whose phenomenology this experiential property contributes to. In Nida-Rümelin's words:<sup>9</sup>

[I]n having an experience we are necessarily aware [...] of ourselves as the one to whom something is phenomenally given [...]. When a tree is phenomenally given to you, you are, in being under that impression, aware of there being something that is phenomenally present to you. This kind of awareness is not something the experience of the tree might have or lack; it is rather part of what it is to have an experience of there being a tree. According to this proposal, having an experience necessarily involves awareness of there being something that is 'given to me'...

Nida-Rümelin (2014, p. 269-70)

<sup>8</sup>See also: "Primitive awareness of instantiating [experiential property] P is *nothing over and above* instantiating P" (Nida-Rümelin 2017, p. 64, my emphasis).

<sup>9</sup>See also Nida-Rümelin (2018, pp. 3379-80).

Thus, by having a conscious experience of a yellow lemon, you are *ipso facto* aware of being phenomenally presented with yellow, that is, aware of being the subject *to whom* a yellow is phenomenally presented. This view might sound puzzling. It is one thing to say that a subject S's instantiation of an experiential property E (e.g., being phenomenally presented with yellow) constitutively contributes to S's overall phenomenology; but it is presumably another to say that S is aware of being the subject whose overall phenomenology E constitutively contributes to. As I have previously argued, having an experience of a yellow lemon certainly puts one in a position rationally to *judge* (and perhaps to *know* in many cases) that one *oneself* is having that experience. But one might wonder in what sense having an experience of a yellow lemon *constitutively* involves being aware of being the subject who is having that experience, or whose overall phenomenology the experience contributes to.

Nida-Rümelin stresses that this awareness of *oneself* is both *pre-reflective*, meaning that it is not a matter of attending to oneself or to one's experience, and *non-conceptual*, meaning that it need not involve the deployment – nor even the possession – of a concept of self (Nida-Rümelin 2017, p. 66). But the question remains: why would all conscious experiences involve a (pre-reflective and non-conceptual) awareness of oneself as the subject of that experience? Consider the experiences of watching a captivating movie, running the final stretch of a marathon, solving a difficult equation, or letting one's mind wander just before falling asleep. Do such experiences really involve an awareness of oneself as an experiencing subject?

One way to answer these questions is to interpret Nida-Rümelin's notion of self-awareness in a very minimal sense, like Legrand's notion of 'pre-reflective self-consciousness'. Indeed, like Legrand, Nida-Rümelin emphasises that "the subject is aware of itself in a way which does not involve its occurrence in its own stream of consciousness" (*ibid.*, p. 67). In Nida-Rümelin's terminology, what it means for an item to "show up in the stream of consciousness" is for it to "[show up] in the totality of what is phenomenally presented to a subject in a given moment" (*ibid.*, p. 63). Thus, she claims the subject is aware of itself in a way that does not involve its being phenomenally presented to itself, as one can be phenomenally presented (for example) with yellow. This suggests

that Nida-Rümelin's notion of pre-reflective self-awareness, like her notion of primitive awareness, can be construed within the scope of foundational claims about consciousness.

However, this picture gets significantly more complicated when one looks in more detail at the three most recent publications in which Nida-Rümelin presents her account of self-awareness (Nida-Rümelin [forthcoming](#); 2017; 2018). Recall that 'experiential properties' are properties that are constitutively such that there is something it is like for a subject to instantiate them. In her (2017), Nida-Rümelin mentions that some (perhaps most) experiential properties are such that "having them consists in being presented with something", which involves a "subject-object structure" that she calls "basic intentionality" (*ibid.*, p. 58). She calls the subset of experiential properties that satisfy this description 'p-experiential properties' (where 'p' stands for 'being presented with something'). On this account, being presented with something (basic intentionally) need not be being *phenomenally* presented with something; for example, something can be "presented to the subject in thought" without there being "phenomenal presence" (*ibid.*, p. 58). So thinking about something is a 'p-experiential property' even though it does not involve being *phenomenally* presented with something, where paradigmatic examples of the latter are being presented with blue, or with a sound (see *ibid.*, p. 55). In summary, the set of experiential properties includes the subset of experiential properties that involve being presented with something (p-experiential properties), and this subset includes yet another subset of experiential properties that involve being *phenomenally* presented with something.

With this distinction in place, Nida-Rümelin states that she will *not* defend the claim that "all experiential properties are such that having them necessarily involves being aware of oneself" (*ibid.*, p. 66); rather, she restricts her claim to *p-experiential properties*: "being presented with something necessarily involves being pre-reflectively and pre-conceptually aware of being the subject to whom something is presented" (*ibid.*, my emphasis). She then goes on to claim that "being so aware of oneself (as the one in the subject position) *does not involve being in any way presented to oneself* ... In other words: pre-reflective self-awareness does not exhibit basic intentionality." (*ibid.*, p. 67, my emphasis). In other words, the relevant kind of self-awareness is not itself a 'p-experiential property', that is, an experiential property that involves being presented with something – this what Nida-Rümelin means

when she claims that the subject does not occur in its own stream of consciousness. But while being aware of oneself in the relevant sense is not a p-experiential property, it is nonetheless “an experiential property since being aware of oneself is *part of what it is like to experience*” (*ibid.*, p. 66, fn. 10, my emphasis). This kind of formulation immediately raises the question of whether Nida-Rümelin’s notion of self-awareness might go *beyond* the foundational claims about consciousness introduced in chapter 1.

Interestingly, Nida-Rümelin considers an objection to her account of self-awareness according to which “[it] is trivially true and does not add anything to the claim that having experiential properties involves being primitively aware of having them” (*ibid.*, p. 68). In other words, the objection points out that it is not clear what exactly pre-reflective self-awareness adds to a subject S’s being primitively aware of S’s instantiation of a p-experiential property (e.g., S’s being primitively aware of what it is like for S to be presented with yellow). In her response to this objection, Nida-Rümelin stresses that her position about self-awareness does add something to her claim about primitive awareness, because “we are self-aware *in a way which is manifest in the way it is like to live through the relevant moment*” (*ibid.*, p. 69, my emphasis).<sup>10</sup> There are two ways to interpret this claim. One could interpret it as stating that pre-reflective self-awareness is ‘phenomenologically manifest’, in the same sense that Kriegel’s inner awareness is taken to be ‘phenomenologically manifest’ – which would suggest that Nida-Rümelin’s position is after all a substantive claim that points to an aspect of phenomenology not already acknowledged by the Nagelian elucidatory account of consciousness. Alternatively, one might interpret this claim as reiterating the now familiar point about the relationship between subjects and experiences being more than *mere instantiation*.

The problem is that Nida-Rümelin is not very explicit about what it means to “be aware of [one]self as an experiencing subject in the phenomenologically manifest way” (*ibid.*, p. 69). She suggests that this notion is illuminated by the distinction between being aware of oneself *as object* and being aware of oneself *as subject*. This distinction originates from

<sup>10</sup>Further down, Nida-Rümelin also writes: “it is a *phenomenological fact* that the subject undergoing such an experience [i.e., an experience consisting in the instantiation of p-experiential properties] is pre-reflectively aware of being the one to whom something is presented” (*ibid.*, p. 69).

a famous passage of Wittgenstein's *Blue Book* in which he identifies two uses of the first-person pronoun in self-ascriptions, 'as object' and 'as subject' respectively (Wittgenstein 1958). According to Wittgenstein, the first-person pronoun is used 'as object' in a self-ascriptive judgement if such judgement is vulnerable to a certain kind of error, namely to the misidentification of the subject of the self-ascription as oneself. In turn, the first-person pronoun is used 'as subject' if the relevant self-ascription is immune to this kind of error. The judgement 'My arm is broken' is an example of the first kind of use ('as object'), while the judgement 'I have toothache' is an example of the second kind of use ('as subject'). Self-ascriptions involving the use of the first-person pronoun 'as subject', in Wittgenstein's sense, are commonly said to be immune to error through misidentification with respect to the first-person (Shoemaker 1968; see also Child 2011): in judging that one has toothache, one cannot misidentify oneself as the subject who has toothache – in other words, one cannot be correct that *someone* has toothache, yet mistaken that *one oneself* has toothache. By contrast, upon seeing a broken arm, one can in some circumstances (e.g., in the immediate aftermath of a car accident) misidentify that arm as one's own – in other words, one can be correct that *someone's* arm is broken, yet mistaken that *one's own* arm is broken.<sup>11</sup>

After Wittgenstein, a number of authors have drawn an analogous distinction between two kinds of self-consciousness: a consciousness of oneself *as object* and a consciousness of oneself *as subject*.<sup>12</sup> This distinction is often fleshed out by saying that consciousness of oneself *as object* is the kind of self-consciousness that grounds self-ascriptions using the first-person pronoun *as object*, while consciousness of oneself *as subject* is the kind of self-consciousness that grounds self-ascriptions using the first-person pronoun *as subject*. For instance, the judgement "My arm is broken" might be grounded in a visual experience of a broken arm, together with a set of beliefs regarding the prior occurrence of a car accident and the anatomical congruency of the broken arm's position with the location of my torso (for example). On this view, the relevant visual experience of a broken arm

<sup>11</sup>See Wittgenstein (1958, p. 67): "It is possible that, say in an accident, I should feel a pain in my arm, see a broken arm at my side, and think it is mine, while in fact it is my neighbour's."

<sup>12</sup>See, for example, Legrand (2007), Nida-Rümelin (2017), D. W. Smith (1986), and Zahavi (2014).

is an instance of consciousness (as) of oneself *as object* because it does not preclude the possibility of misidentifying the arm that one sees as one's own. By contrast, the judgement "I have toothache" is normally grounded in the subjective experience of toothache, whose occurrence does preclude the possibility of misidentifying it as one's own. If any conscious experience *e* can in principle ground the self-ascriptive judgement "I am undergoing *e*", and if such judgement formed on the basis of undergoing *e* makes use of the first-person pronoun *as subject* (i.e., it is immune to error through misidentification with respect to the first-person), then any conscious experience should constitutively involve consciousness of oneself *as subject* in the relevant sense.

However, something more needs to be said to understand what being conscious of oneself *as subject* really involves. Proponents of this notion generally emphasize that self-consciousness *as subject* involves a special connection to the self that guarantees the immunity to error through misidentification of self-ascriptive judgements formed on the basis of experiences involving such self-consciousness. How are we to understand this special connection? One might say that being conscious of oneself *as subject* "does not involve being presented to oneself as an object" (Shoemaker 1968, p. 563) – but such formulation has an air of tautology.<sup>13</sup> Nida-Rümelin does say a bit more about the relevant notion: "[w]hen a subject instantiates a p-experiential property, then it is aware of being in the subject position of basic intentionality" (Nida-Rümelin 2017, p. 67). In other words, in being presented with something, the subject is aware of being in the subject position of the 'subject-object structure' that characterises the event of being presented with something; but such self-awareness is not itself an instance of being presented *with oneself* ('as an object'), although it is an experiential property, and in that respect constitutively contributes to what it is like for the subject.

In light of this analysis, we can see that Nida-Rümelin's account of self-awareness in her (2017) seems to go beyond what she says in her (2018), namely that "[p]rimitive

<sup>13</sup>See also Nida-Rümelin (2017, p. 73): "in pre-reflective self-awareness... the subject is not given to itself as an object"; and Legrand (2007, p. 587): "consciousness of the self-as-*subject* [is] a non-objectifying form of self-consciousness where the self is not an intentional object of consciousness detached from the subject of consciousness".

awareness of experiential properties therefore comes along with some form of awareness of oneself [such that] phenomenal consciousness and pre-reflective self-awareness are just two sides of one and the same ‘phenomenon’” (2018, p. 3380). In her (2017), Nida-Rümelin claims that (a) only primitive awareness of *p-experiential properties*, rather than primitive awareness of all experiential properties, involves pre-reflective self-awareness, and (b) such pre-reflective self-awareness is *phenomenologically manifest* in what it is like to instantiate p-experiential properties. This is presumably not equivalent to the claim that phenomenal consciousness and pre-reflective self-awareness are just two aspects of the same phenomenon.

In her (forthcoming), Nida-Rümelin is more explicit. She distinguishes between three meanings of the phrase ‘subjective character’:

- An experience *e* of a subject S has [subjective character]<sub>1</sub> if and only if *e* consists in the instantiation of experiential properties by S.
- An experience *e* of a subject S has [subjective character]<sub>2</sub> if and only if S is primitively aware of having the experiential properties whose instantiation *e* consists in.
- An experience *e* of a subject S has [subjective character]<sub>3</sub> if and only if S is pre-reflectively aware of S [SELF] as being the subject to whom something is presented.

Nida-Rümelin clarifies that on her view, [subjective character]<sub>1</sub> and [subjective character]<sub>2</sub> really are like two faces of the same phenomenon. This is consistent with her definition of primitive awareness: being primitively aware of an experiential property simply is to instantiate that property. However, Nida-Rümelin also clearly states that [subjective character]<sub>3</sub> is *not* equivalent to either [subjective character]<sub>1</sub> or [subjective character]<sub>2</sub>: it points to some *further* aspect of phenomenology:

It is a substantial and interesting fact about the nature of experiential properties that having them coincides with being primitively aware of having them. For an experience to have [subjective character]<sub>3</sub>, by contrast, is a different fact. It is not the same, or so the discussion presented suggests, as the fact which renders attributions of [subjective character]<sub>1</sub> and [subjective character]<sub>2</sub> correct. Contrary to the latter two senses of having subjective character, for an experience to have [subjective character]<sub>3</sub> is for it to come along with a certain phenomenal aspect: the subject is aware of itself in a *phenomenally manifest manner*. To say that an experience has [subjective

character]<sub>3</sub> is to make a statement about what it is like to undergo the experience. To say of an event that it has [subjective character]<sub>1</sub> or [subjective character]<sub>2</sub> is to say of it that it belongs to the ontological category of experiences.

Nida-Rümelin ([forthcoming](#))

In line with what she says in her ([2017](#)), Nida-Rümelin adds that she is *not* committed to the claim that “there is a *necessary* connection between experiencing and being pre-reflectively aware of oneself” (*ibid.*, my emphasis). Rather, she tentatively endorses the claim that pre-reflective self-awareness is “present in all mature human experience” (*ibid.*). Thus, she ultimately refrains from endorsing a claim like (CSC<sub>E</sub>) if ‘pre-reflective self-awareness’ is understood as a notion that goes *beyond* primitive awareness. But she also adds in a footnote that there is a notion of awareness of *oneself* that does *not* go beyond primitive awareness:

One may object that being aware of having an experience involves being aware of oneself just like any awareness of an instantiation of a property involves being aware of the object instantiating it. I agree that there is a sense in which primitive awareness involves awareness of oneself but I deny that it implies being aware of oneself in a phenomenally manifest manner *as* the subject undergoing the experience.

Nida-Rümelin ([forthcoming](#), fn. 27)

Consequently, it might be possible to reconcile what Nida-Rümelin says in her different papers if one interprets her position as the combination of *two* distinct claims: (i) a minimal version of (CSC<sub>E</sub>), according to which, instantiating an experiential property involves being aware of *oneself*, but not in a phenomenologically manifest manner; (ii) a stronger claim according to which all mature human experience (according to her [forthcoming](#)), or perhaps all instantiations of p-experiential properties (according to her [2017](#)), involve a phenomenologically manifest awareness of oneself as the subject undergoing the experience. The second claim is presumably not a version of (CSC<sub>E</sub>), in so far as it might not apply to *all* conscious experiences. The first claim, by contrast, is a version of (CSC<sub>E</sub>).

Let us take stock. I have argued that there are two versions of (CSC<sub>E</sub>) that do not seem to go beyond the foundational claims discussed in chapter 1. For example, the versions of (CSC<sub>E</sub>) defended by Legrand and Nida-Rümelin seem to fit within that category. While Nida-Rümelin also articulates a seemingly stronger notion of self-awareness, she denies that this stronger notion is *constitutive* of consciousness. The upshot of this analysis of

‘minimal’ interpretations of (CSC<sub>E</sub>) is rather similar to what I concluded from my discussion of ‘minimal’ interpretations of (CSC<sub>NE</sub>). Glossing the idea that experiences constitutively contribute to their subject’s overall phenomenology in terms of consciousness or awareness of *oneself* is illuminating in so far as this notion can be constructed within the scope of the foundational claims about consciousness discussed in chapter 1. Nonetheless, the formulation ‘consciousness (awareness) of oneself’ is quite suggestive, and can easily be mistaken for a more substantive notion; in that respect, it is perhaps safer to use some alternative terminology, such as the terminology I have used in chapter 1.

### 4.3 Substantive constitutive claims

I have argued so far that there is a notion of egological self-consciousness, or consciousness of *oneself*, that – upon close examination – does *not* actually go beyond foundational claims about consciousness. This notion is Legrand’s and Nida-Rümelin’s minimal notion of pre-reflective self-consciousness, according to which being in a conscious mental state M constitutively involves being conscious (or aware) of being the subject of M, in the minimal sense that M constitutively contributes to one’s overall phenomenology. It is also *prima facie* equivalent to Zahavi’s minimal notion of pre-reflective self-consciousness.

I have suggested that there is a stronger – and perhaps more idiomatic – egological notion of self-consciousness that could be employed in a less minimal interpretation of (CSC<sub>E</sub>). According to this stronger notion, being self-conscious is a matter of being conscious of oneself as oneself in such a way that one’s consciousness of oneself as oneself is a component of one’s overall phenomenology. For clarity, we may call this stronger egological notion of self-consciousness *phenomenal self-consciousness*, to distinguish it from the minimal egological notion of pre-reflective self-consciousness:

(PSC) A subject S is *phenomenally self-conscious* at *t* if and only if (a) S is conscious of S [SELF] at *t*, and (b) S’s being conscious of S [SELF] at *t* makes a determinate constitutive contribution to S’s overall phenomenology at *t*.

Thus, phenomenal self-consciousness does not merely point to the fact that there is something at all, rather than nothing at all, that it is like for the subject; it also points

to a *specific* aspect of what it is like for the subject, namely what we might call on first approximation a *sense of self*. In Nida-Rümelin's terminology, one might say that being phenomenally self-conscious is to be phenomenally presented with oneself – in such a way that, as she would also put it, the self occurs in one's stream of consciousness.

If one interprets the egological notion of self-consciousness at work in (CSC<sub>E</sub>) as phenomenal self-consciousness, then (CSC<sub>E</sub>) entails that whenever one is conscious, there is a component of one's overall phenomenology that corresponds to one's consciousness of *oneself as oneself* – perhaps some form of “phenomenology of self” that, as Chalmers (1996) puts it, is present in every conscious experience like “a kind of background hum... that is somehow fundamental to consciousness” (p. 10). This interpretation of (CSC<sub>E</sub>) clearly goes beyond the foundational claims discussed in chapter 1, and thus requires independent motivation.

As we have seen, Kriegel (2009) is tempted by this substantive version of (CSC<sub>E</sub>), although he falls short of endorsing it as a claim about what is *constitutive* of consciousness. He suggests instead that consciousness or awareness of *oneself as oneself* – and perhaps as the sort of thing that has one's autobiographical memories – is a ubiquitous feature of the conscious lives of *normal human adults*:<sup>14</sup>

For my part, I find myself attracted to the view that pre-reflective self-consciousness in normal human adults is a genuine form of awareness [of] oneself as (potentially autobiographical) subject... I am much less confident of this claim, however, than of the claim that peripheral inner awareness is a constitutive aspect of phenomenology.

Kriegel (2009, p. 179, fn. 17)

Kriegel is not isolated in acknowledging some sympathy for this kind of view. Thus, Block (1995) seems tempted by a somewhat similar view when he writes that “[phenomenally] conscious states often seem to have a ‘me-ishness’ about them; the phenomenal content often represents the state as a ‘state of me’” (p. 235). Like Kriegel, Block is careful not to claim that the relevant ‘me-ishness’ is *constitutive* of consciousness. As I noted in the general introduction, Chalmers himself mentions that “there seems to be *something* to

<sup>14</sup>See also Kriegel (2009, pp. 42-3).

the [ubiquitous] phenomenology of self, even if it is very hard to pin down” (1996, p. 10); but he does not elaborate on how one might defend this view.

It proves difficult to find more confident endorsements of the non-minimal interpretation of (CSC<sub>E</sub>). Quite often, passages that *seem* at first to endorse a non-minimal reading of (CSC<sub>E</sub>) end up equivocating between this reading and the more minimal reading interrogated in the previous section. An early example of this trend can be found in an interesting article on consciousness and self-consciousness published by William Henry Scott in *Mind* in 1918:

It is certain that there is a ‘self-quality’ in all experience, even the most primary. Self is always present, whether consciously or not, and its presence necessarily imparts a distinctive character to the experience. It is like a condiment, which affects the taste of our food though we take no distinct notice of it. Experience is flavored through and through with a self-quality even when we do not recognize it. My experience is *mine* whether I recognize the relation or not. ‘The baby new to earth and sky has never thought that “this is I”’; yet such experience as he has is *his*, and all its meaning and value to him grow out of that fact. But this self-quality is not what I mean by self-consciousness. It is rather a sign or prophecy of what may be. Self-consciousness is a *recognition* of self as being conscious. Even so, it is of many degrees. It may be an almost unobserved element, barely emerging into light and deeply overshadowed by the consciousness of the object.

Scott (1918, p. 10)

This passage initially seems to defend the strong claim that every experience necessarily involves a sense of self that is part of what it is like to have it (i.e., what I have called phenomenal self-consciousness). But then Scott goes on to characterise this as the fact that my experience is *mine*, or that a baby’s experience is *his* – which does not entail that one is conscious of oneself as the subject of one’s experiences. Moreover, he qualifies the claim that “[the] Self is always present” in experience by adding “consciously or not”. The idea that the self can be non-consciously present in experience might simply refer to the fact that all experiences are undergone – enjoyed or endured – *by a subject*, whether or not this subject is conscious *of herself*. Scott also contrasts the ‘self-quality’ of experience with ‘self-consciousness’ construed as a *recognition* of oneself. The term ‘recognition’ might evoke a rather sophisticated and salient process that would leave room for ‘self-quality’ being, after all, a simpler, pre-reflective and non-recognitional sense of self. But he goes on to say that self-consciousness itself comes in many degrees, and can be barely noticeable at all, which

suggests on the contrary that the relevant ‘recognition’ of self need not be so sophisticated and salient. In any case, Scott asserts his position as evident (“it is certain...”), and does not provide a positive argument for it.

One might read into another passage cited by Nathan (1997) – seemingly to endorse it – a bolder yet even less explicit formulation of a strong interpretation of (CSC<sub>E</sub>):

I consider... my consciousness and feeling of myself, that taste of myself [to be] more distinctive than the taste of ale or alum, more distinctive than the smell of walnut-leaf or camphor.

Hopkins (1959, p. 123)

However, this passage is not taken from a philosophical work but from a sermon; and Hopkins does not elaborate on whether the ‘feeling of myself’ is constitutively present in all conscious experiences, let alone provide an argument to support this controversial claim.

As these examples show, it is not easy to find clear positive arguments in favour of (CSC<sub>E</sub>) that go beyond the minimal interpretation put forward by Legrand and Nida-Rümelin. Many authors are happy to acknowledge a general *inclination* to endorse a claim in the vicinity of (CSC<sub>E</sub>), according to which most conscious experiences – perhaps only in normal human adults – involve phenomenal self-consciousness. But far fewer authors are willing to extend this claim to *all* conscious experience, and to characterise phenomenal self-consciousness as a *constitutive* feature of consciousness. There is, in my opinion, a good reason for this: there is simply no good argument for a substantive reading of (CSC<sub>E</sub>). Upon elucidating the notion of consciousness as we have done in chapter 1, it becomes clear that being conscious at all is *not* a matter of being conscious *of oneself as oneself*, unless the relevant notion of self-consciousness is intended to point only to the fact that one’s experiences constitutively contribute to one’s overall phenomenology. If one were to claim unambiguously and confidently that being conscious constitutively involves being conscious *of oneself* in a stronger sense – i.e., in the sense that what it is like to be in *any* conscious experiences necessarily involves phenomenal self-consciousness –, one would have to provide independent support for this claim beyond what the Nagelian account of consciousness can offer. I must confess that I cannot see how one might convincingly provide such motivation.

Let us recapitulate what has been argued in this chapter so far. I suggested that formulations of  $(CSC_E)$  can be in principle divided into two categories. In the first category are ‘minimal’ interpretations of  $(CSC_E)$  that do not actually go beyond the foundational claims of chapter 1. In the second category are more substantive interpretations of  $(CSC_E)$  that go beyond such foundational claims. I argued that claims that belong to the first category – such as Legrand’s and Nida-Rümelin’s versions of  $(CSC_E)$  – are plausible, because they plausibly follow from the Nagelian elucidatory account of consciousness. I subsequently argued that it is difficult to find clear formulations of claims that belong to the second category, and even more difficult to find clear positive arguments for such claims. This should not be surprising, for the more substantive interpretation of  $(CSC_E)$  is particularly implausible. Being conscious at all necessarily involves being in a conscious mental state that constitutively contributes to one’s overall phenomenology; but a theoretical elucidation of the notion of consciousness certainly does not suggest that being conscious should also necessarily involve being phenomenally self-conscious, where this is taken to be a *further* fact about consciousness.

The outcome of the long discussion that runs from the previous chapter to the present section is that we should abandon what I have provocatively called the myth of constitutive self-consciousness. If the relevant notion of self-consciousness points to aspects of phenomenology that are not acknowledged by foundational claims about consciousness, then it is incorrect to claim that self-consciousness is constitutive of consciousness. If, on the other hand, the relevant notion of self-consciousness is intended in a minimal sense that can be constructed from the resources provided by the Nagelian elucidatory account of consciousness, then it is plausible and potentially illuminating; but it is preferable to gloss such a notion without the terminology of ‘self-consciousness’ that evokes more substantive and determinate aspects of phenomenology.

#### 4.4 Non-constitutive substantive claims

I have argued so far that there is no compelling argument to support the claim that self-consciousness is constitutive of consciousness, whether the former is construed as consciousness *of one’s experience* or consciousness *of oneself* – unless these notions are not

intended in a minimal sense. Given the Nagelian definition of consciousness, it is not conceptually possible for a subject to have a conscious experience that does not constitutively contribute to the subject's overall phenomenology. However, it is at least conceptually possible for a subject to have a conscious experience without being conscious *of that experience* or conscious *of herself*, in so far as either of these notions go beyond the Nagelian elucidation of consciousness.

Importantly, phenomenal features that are not constitutive of consciousness might still play important roles in the mental life of humans and non-human animals. Visual phenomenology, for example, is clearly not constitutive of consciousness, since blind individuals are no less conscious than the rest of the population. Yet such phenomenology is pervasive in the experience of sighted individuals, and plays a major role in guiding their behaviour. The same might be true of self-consciousness; perhaps the vast majority of conscious experiences do involve some form of self-consciousness, even if there are some cases in which it can be exceptionally missing. Whether or not this is the case is an empirical question, and my arguments against (CSC<sub>E</sub>) do not carry any implications regarding the prevalence of egological self-consciousness in conscious experience.

I noted earlier that a number of authors are generally *inclined* to agree that many conscious experience involve phenomenal self-consciousness, namely the egological notion of self-consciousness that does not point to aspects of phenomenology already acknowledged by the Nagelian account. While few philosophers have endorsed the strong version of (CSC<sub>E</sub>) according to which phenomenal self-consciousness is *constitutive* of consciousness (an aspect of all conscious mental states), many seem to find it plausible that *ordinary* conscious experience involves phenomenal self-consciousness. Let us call this the *typicalist claim* about phenomenal self-consciousness, or (TSC) for short:

(TSC) *In ordinary circumstances*, if a subject S is conscious, (a) S is conscious of S [SELF], and (b) S's being conscious of S [SELF] makes a determinate constitutive contribution to S's overall phenomenology at *t*.

In this formulation of (TSC), I purposefully leave open what one might build into the reference to 'ordinary circumstances'. One might interpret this qualification as a restriction

on the type of subject to which (TSC) applies. For example, Kriegel seems to suggest that the relevant form of egological self-consciousness is a ubiquitous feature of the conscious lives of *normal* (e.g., healthy and ‘neurotypical’) *adult human beings*. Alternatively, one might interpret ‘ordinary circumstances’ as a restriction on the frequency at which (TSC) applies. For example, Block seems to suggest that some form of self-consciousness is *often* or *typically* part of one’s phenomenology, but not *always*. One might specifically claim that (TSC) applies in the ordinary wakeful state, but not in altered states of consciousness (such as during dreams, meditation or drug-induced states); or one might claim that (TSC) does not apply in specific pathological conditions (such as psychosis or depersonalisation disorder).

While I do not accept (CSC<sub>E</sub>) on a substantive interpretation, I think (TSC) is antecedently plausible. Note that unlike (CSC<sub>E</sub>), (TSC) is a contingent empirical claim: it is a claim about what empirically applies *in some circumstances* in which a subject is conscious, not a claim about what necessarily applies *in all circumstances* in which a subject is conscious. As such, (TSC) can and should be supported in principle by an examination of phenomenology – an examination of what it is like for a subject *in ordinary circumstances*. In other words, if (TSC) is true, then an examination of the overall phenomenology of conscious subjects *in ordinary circumstances* (whatever may be intended by this qualification) should in principle confirm that some form of consciousness *of oneself as oneself* is part of that phenomenology. Furthermore, if one’s circumstance at some time *t* is included within the subset of ‘ordinary circumstances’ in which (TSC) holds, then it should in principle be possible to assess (TSC) by introspecting one’s own phenomenology at *t*. Indeed, introspection is presumably the mechanism through which Block, Chalmers and Kriegel find support for (TSC), as their choice of terminology reveals: “conscious states often *seem* to have a ‘me-ishness’ about them” (Block 1995, p. 235, my emphasis); “there *seems* to be something to the phenomenology of self” (Chalmers 1996, p. 10, my emphasis); “my current experience’s pre-reflective self-consciousness *strikes me* as egological” (Kriegel 2009, p. 177, my emphasis).

However, there has been no shortage of scepticism regarding the idea that we can find anything like a sense of self in experience, since Hume’s infamous remark:

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

Hume (1978, p. 252)

Thus, a claim like (TSC) is likely to generate an introspective disagreement, that is, a disagreement as to whether introspection does reveal, *in the relevant ordinary circumstances*, that one's overall phenomenology includes phenomenal self-consciousness. Introspective disagreements are difficult to arbitrate, as they often lead to seemingly intractable disputes in which it is hard for both camps to find common ground (Bayne and Spener 2010). In particular, it is possible that the disagreement about (TSC) might hinge at least in part on terminological variation. Indeed, the two sides of the disagreement might lack a common understanding of what they are supposed to notice in their experience, although they could perhaps come to an agreement if the target feature was specified in more detail. Consider, for example, Prinz' sceptical stance on the possibility of experiencing the self in isolation:

Among the various phenomenal qualities that make up an experience, there is none that can be characterized as an experience of the self... There is no phenomenal I. If I wait for myself to appear in experience, I will never arrive.<sup>15</sup>

Prinz (2012, pp. 214-240)

On this neo-Humean view, one might say (paraphrasing Nida-Rümelin) that there is no *phenomenal presentation of the self as oneself*. In other words, there is no distinctive experience of the self – no qualitative property of 'selfhood', as it were – among the various phenomenal qualities that show up in one's stream of consciousness. But perhaps Prinz' notion of an 'experience of the self' or a 'phenomenal I' is not specific enough to allow both sides of the disagreement to agree on what they can or cannot notice in their phenomenology.

Consider, by analogy with (TSC), a typicalist claim about 'world-consciousness', or (TWC) for short:

(TWC) *In ordinary circumstances*, if a subject S is conscious, then (a) S is conscious of *the world* and (b) S's being conscious of *of the world* makes a determinate constitutive contribution to S's overall phenomenology.

<sup>15</sup>See Howell (2006) and Dainton (2008) for similar considerations.

The relevant notion of ‘world-consciousness’ or ‘consciousness of the world’ might seem *prima facie* mysterious. Paraphrasing Hume’s comment on the sense of self, one might take the following sceptical stance on the matter:

For my part, when I enter most intimately into what I call *the world*, I always stumble on some particular perception or other. I can never catch *the world* at any time without a perception, and can never observe anything but the perception.

However, this sceptical stance fails to account for the possibility that the notion of *world-consciousness* might refer generically to any conscious experience of a worldly object – any instance in which one is conscious of the external world in one way or another – rather than something over and above any such experience. Thus, being conscious of a table (*table-consciousness*, as it were) would be a particular instance of *world-consciousness*. More specifically, *world-consciousness* would be the determinable of which *table-consciousness* is a determinate, much as *being coloured* is a determinable of which *being red* is a determinate. Moreover, *world-consciousness* might have a number of determinates – such as *chair-consciousness*, *lamp-consciousness*, etc. – each of which is a different instance of being conscious of a worldly object. Now, if subjects facing a table were asked whether their experience involves *table-consciousness* defined as consciousness of a table, it is unlikely that this question would yield any introspective disagreement.

In this toy example, the initial disagreement about the existence of *world-consciousness* in ordinary experience was simply an unfortunate consequence of the ambiguity of the notion. Disambiguating the notion involves specifying that it refers to the determinable of which *table-consciousness*, *chair-consciousness*, *lamp-consciousness*, etc. are determinates, rather than consciousness of something over and above the chair, the table, the lamp, etc. Once the notion has been properly disambiguated, (TWC) can be broken down into several more specific claims that are more likely to generate agreement.

Similarly, one could consider the relevant notion of self-consciousness in (TSC) as a determinable of which there are several determinates. This would involve distinguishing between several candidate forms of phenomenal self-consciousness – several ways of being conscious of oneself as oneself that make some *determinate* contribution to one’s phenomenology, and thus go beyond the ‘minimal’ interpretation of egological

self-consciousness. Whether such forms of phenomenal self-consciousness *actually* occur in experience, and in what circumstances or with what frequency, are empirical questions that require us to examine evidence about phenomenology. I will explore these questions in the second part of the present thesis, by considering three candidate determinates of phenomenal self-consciousness.

The most obvious way in which one can be phenomenally self-conscious is to engage in conscious thinking about oneself as oneself, or *cognitive self-consciousness*. In chapter 5, I will argue that cognitive self-consciousness is a manifest aspect of one's phenomenology, and often occurs in ordinary conscious experience, although it is certainly not pervasive. But engaging in conscious thinking about oneself as oneself is not the only determinate of which phenomenal self-consciousness is a determinable. In chapters 6 and 7, I will focus on two other candidate determinates of phenomenal self-consciousness that play a particularly important role in our conscious mental lives. The first of these, considered in chapter 6, is the consciousness of oneself as the bodily subject whose body part one feels a sensation in or on, or *bodily self-consciousness*. I will argue that bodily self-consciousness is a relatively pervasive feature of ordinary conscious experience, and is not merely a species of cognitive self-consciousness, but a genuinely distinct form of phenomenal self-consciousness. In chapter 7, I will turn to another candidate determinate of phenomenal self-consciousness, namely the consciousness of oneself as being located at the origin of the spatial perspective of one's perceptual experience, or *spatial self-consciousness*. I will argue that spatial self-consciousness is ubiquitous in ordinary circumstances, and is not merely a species of cognitive self-consciousness either, but yet another genuinely distinct form of phenomenal self-consciousness. Taken together, chapters 5, 6, 7 thus constitute a defence of what we might call a *pluralist account* of phenomenal self-consciousness, according to which there are several ways in which one can be phenomenally self-conscious.

One noteworthy consequence of my focus on (TSC) and phenomenal self-consciousness in the second part of this thesis is that the *non-egological* notion of self-consciousness (consciousness of consciousness itself, or consciousness of one's experience) will fall out of the picture. There are several reasons for this. First, the idea that self-consciousness might be pervasive *in ordinary circumstances* yet not be *constitutive* of

consciousness has little appeal if one construes the relevant notion as consciousness *of one's experience*. Indeed, it is difficult to find authors who argue that we are conscious of our experiences *most of the time*, but not *always*. Second, as I have already noted in chapter 2, the most natural understanding of the notion of 'self-consciousness' is the egological notion of self-consciousness, and more specifically the notion of phenomenal self-consciousness. The non-egological notion of self-consciousness is a rather unusual construal of the term 'self-consciousness' that was cemented in philosophical parlance by the phenomenological tradition. Third, if one were to examine a claim analogous to (TSC) in which the relevant notion of self-consciousness is non-egological, one would most likely end up interpreting this notion as a form of *introspection* (since being in a mental state does not *constitutively* involve introspecting it); the resulting discussion would then focus on the nature and prevalence of introspection in ordinary experience. Quite a lot has been written already about introspection; rather less has been written about whether there might be distinct ways in which one may be phenomenally self-conscious. For these reasons, I will now leave non-egological self-consciousness to the side – as well as the minimal interpretation of egological self-consciousness as 'pre-reflective self-consciousness' – to investigate the nature and prevalence of substantive phenomenal self-consciousness in the real world.

## II

# SELF-CONSCIOUSNESS IN THE REAL WORLD

# 5

## Cognitive Self-Consciousness

It is good to think about oneself, but one should not concern oneself too much with it.

Arnauld (1660/1775, p. 217, my translation)

In the first part of this thesis, I considered whether self-consciousness is *constitutive* of consciousness. This question has led me to distinguish a number of different notions of self-consciousness, starting with a broad distinction between (a) self-consciousness in the *non-egological* sense (consciousness of *one's experience*) and (b) self-consciousness in the *egological* sense (consciousness of *oneself*). For each of these broad notions of self-consciousness, I have distinguished between (i) more specific 'minimal' or deflationary notions that point to, and further elucidate, aspects of phenomenology already acknowledged by a Nagelian elucidatory account of consciousness (in particular, what I have called 'phenomenality', the second-order property of there being something rather than nothing that it is like to be in a mental state), and (ii) more specific 'thick' or inflationary notions that point to some further and determinate aspect of phenomenology – determinate components of what it is like to undergo an experience. This double conceptual distinction is summarised in figure 5.1.

At the end of chapter 4, I focused more closely on the 'thick' egological notion of self-consciousness that I have called 'phenomenal self-consciousness', defined as follows:

(PSC) A subject *S* is *phenomenally self-conscious* at *t* if and only if (a) *S* is conscious of *S* [SELF] at *t*, and (b) *S*'s being conscious of *S* [SELF] at *t* makes a determinate constitutive contribution to *S*'s overall phenomenology at *t*.

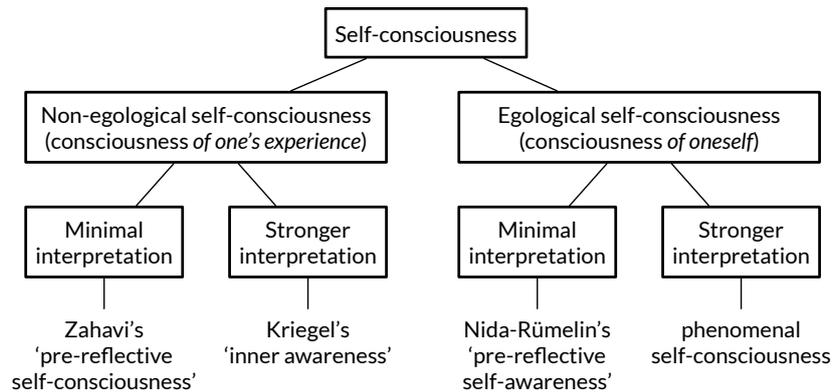


Figure 5.1: A disambiguation of 'self-consciousness'.

I observed that a number of philosophers seem tempted by (TSC), the claim that in some set of ordinary circumstances to be further specified, a subject *S* is phenomenally self-conscious if she is conscious at all:

(TSC) *In ordinary circumstances*, if a subject *S* is conscious, (a) *S* is conscious of *S* [SELF], and (b) *S*'s being conscious of *S* [SELF] makes a determinate constitutive contribution to *S*'s overall phenomenology at *t*.

However, I also remarked that this claim is not easy to assess if the notion of phenomenal self-consciousness is left indeterminate, for there might several ways in which one can be conscious of oneself as oneself such that this consciousness of oneself as oneself is a determinate component of one's overall phenomenology. The most obvious candidate determinate of which phenomenal self-consciousness is a determinable is *conscious thinking about oneself as oneself*. In fact, a number of authors seem to assimilate the ability to be self-conscious in general with the capacity to think of oneself as oneself.<sup>1</sup> In this chapter, I will motivate the weaker claim that engaging in conscious thinking about oneself as oneself is *ipso facto* being phenomenally self-conscious.<sup>2</sup> Furthermore, I will argue that if conscious thinking about oneself as oneself is the only determinate of which phenomenal self-consciousness is a determinable, then (TSC) is empirically very implausible.

<sup>1</sup>See, for example, Rödl (2007, p. 1), Musholt (2015, p. 2), Baker (2000, p. 60), and J. Smith (2017).

<sup>2</sup>I shall argue in subsequent chapters that engaging in conscious thinking about oneself as oneself is not the *only* way in which one can be phenomenally self-conscious.

In §5.1, I introduce the general notion of *de se* thoughts, or thoughts about oneself as oneself, through a series of conceptual distinctions. In §5.2, I argue that a subject engaged in conscious *de se* thinking is *ipso facto* phenomenally self-conscious. In §5.3, I argue that conscious *de se* thinking is not ubiquitous in all conscious experiences. Finally, in §5.4, I discuss empirical evidence that conscious *de se* thinking is not even ubiquitous in the ordinary wakeful consciousness of normal adult human beings.

### 5.1 Thinking about oneself as oneself

Thoughts are structured by intentional content – they are *about* something. Some thoughts are *de dicto* (‘about what is said’ in Latin), such as the thought “The highest mountain on earth rises over 8,000 meters”. This kind of thought is not ‘directly’ about a particular object, but rather about whatever happens to satisfy the descriptive condition “the highest mountain on earth”. Other thoughts are singular or *de re* (‘about the thing’ in Latin): they are about a specific object – for example, the thought “*That mountain* rises over 8,000 meters”, as thought by an alpinist looking at mount Everest.

Thoughts about oneself can be *de dicto* or *de re*. Suppose that I am the tallest man in the room. If I think “The tallest man in the room is over six feet tall”, this *de dicto* thought will be *de facto* (‘as a matter of fact’) about myself – whether or not I happen to know that I am the tallest man in the room. In turn, if I catch a glimpse of my reflection in a mirror, I might think “*That man* is over six feet tall”; again, this *de re* thought would be *de facto* about myself, whether or not I happen to realize that the man in the mirror is *myself*.

We might refer to thoughts about oneself in general – whether such thoughts are *de dicto* or *de re* – as *de se* thoughts loosely speaking (where *de se* means ‘about oneself’ in Latin). As we will see, the meaning of the phrase ‘*de se* thought’ is rather inconsistent in the literature. It is common to consider *de se* thought as a subspecies of *de re* thought; but given that ‘*de se*’ literally means ‘about oneself’, and that *de dicto* thoughts can also be about oneself, it seems more intuitive to call all thoughts about oneself *de se* thoughts, at least in first approximation. By contrast with this general notion of *de se* thought, we might refer to thoughts that are *not* about oneself as *de mundi* thoughts (or thoughts ‘about the world’). As we have seen, *de mundi* thoughts can themselves be either *de dicto* (e.g.,

“The highest mountain on earth rises over 8,000 meters”) or *de re* (e.g., “*That mountain rises over 8,000 meters*”).

There is a further distinction to be drawn within the category of thoughts that are both *de re* and *de se*, between thoughts that are about oneself *de facto* (‘as a matter of fact’, or merely accidentally), and thoughts that are about oneself *de jure* (‘by law’, or non-accidentally). Only the latter deserve to be called thoughts about oneself *as oneself*. This distinction is a special case of the general distinction discussed in chapter 2 between mental states that are accidentally about oneself and those that are non-accidentally about oneself. Thoughts that are about oneself *de jure* form a subset of *de re* thoughts about oneself, in so far as they refer to a singular individual – *myself*. To illustrate this idea, consider Perry (1979)’s famous shopper example (see also Lewis 1979):

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

Perry (1979, p. 33)

Prior to his epiphany, the protagonist of this story may have the thought “The shopper with the torn sack is making a mess”. This is a *de dicto* thought that is accidentally about oneself, because the thinker himself happens to fit – unbeknownst to him – the descriptive condition “the shopper with the torn sack”. By contrast, after realizing that he *himself* is responsible for the trail of sugar on the floor, the shopper might entertain the thought “I am making a mess”. This is not only a *de re* thought, in so far as it is directly about a particular individual, but it is also specifically a *de se* thought that is *de jure*, or non-accidentally, about oneself. The resulting taxonomy of thoughts about oneself, including the special case of *de se* thoughts that are non-accidentally about oneself, is represented on figure 5.2.

For the sake of simplicity, I will henceforth refer to thoughts that are non-accidentally about oneself (namely, thoughts that are conjointly *de re*, *de se* and *de jure* in the above taxonomy) simply as *de se* thoughts. *De se* thoughts, in this strict sense, form a special case

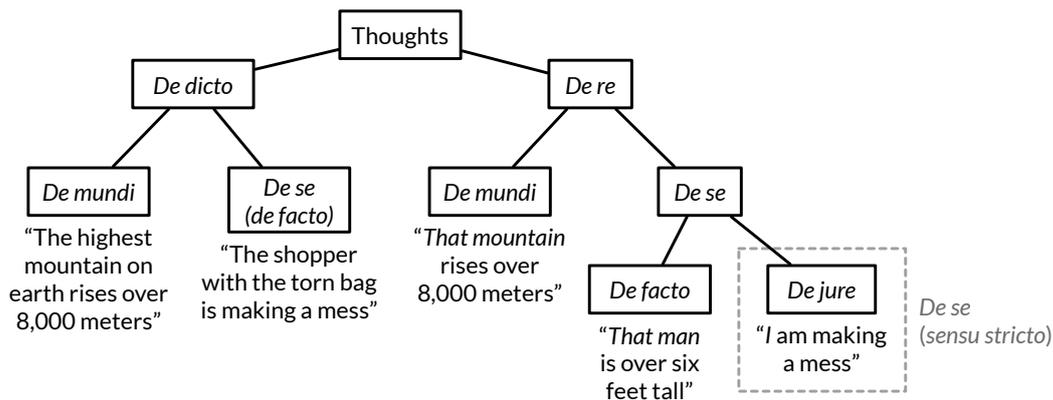


Figure 5.2: A taxonomy of thoughts about oneself.

of ‘*de se* thoughts’ broadly construed as thoughts about oneself. From now on, all mentions of *de se* thought will refer to *de se* thought in the strict sense.<sup>3</sup>

Generally, *de se* thoughts are thoughts that make use of the first-person concept (or concept of self), and are suitably expressed with the first-person pronoun. Indeed, the use of the first-person concept secures the reference to oneself *as oneself*, as opposed to an accidental reference to oneself under a descriptive condition (e.g., “the tallest man in the room”) or with a third-person demonstrative concept (e.g., “that man” in reference to what one sees in the mirror). This is the case because the first-person concept is individuated by the following subject-reflexive condition: in any mental state whose intentional content involves the first-person concept, the latter refers *de jure* to the subject of that mental state.<sup>4</sup> This condition is easily transposed to the specific case of thoughts: in any thought whose intentional content involves the first-person concept, the latter refers *de jure* to the thinker of that thought. Thus, *de se* thoughts are thoughts whose content involves the first-person concept and refers *de jure* to the thinker.

There is a final conceptual distinction to be drawn between *conscious* and *non-conscious* thinking. It is at least conceptually possible that one might think *that p* without there being

<sup>3</sup>Note that some authors consider that only *de re* thought can be *de se*, but refer to all *de re* thoughts that are about oneself (accidentally or not) as *de se* thoughts. See for example Recanati (2009), who distinguishes between *de se* thoughts in general (that may or may not be accidentally about oneself) and ‘first-person thoughts’ that are non-accidentally or *de jure* about oneself. In what follows, I will reserve the notion of *de se* thought for what Recanati calls ‘first-person thoughts’.

<sup>4</sup>See Peacocke (2014, pp. 10-14).

anything at all that it is like for one to think *that p*. Perhaps such instances of non-conscious thinking can and do have a causal influence on behaviour. Upon arriving at a stuffy soirée, for example, I might non-consciously think that I am underdressed, and this might cause me to avoid eye contact with other guests, although I might not realise what caused this change in my behaviour, as there was (constitutively) nothing that it was like for me, for me to think that I am underdressed. By contrast, there must be (constitutively) something that it is like to engage in *conscious* thinking. In so far as it is possible to think *consciously* or *non-consciously*, it must also be possible, specifically, to think *consciously* or *non-consciously* about oneself as oneself. (In the above example, my thinking that *I* am underdressed is an instance of *non-conscious de se* thinking.)

It is worth mentioning that there are various modes of thinking that may involve different types of propositional attitudes, such as consciously (or non-consciously) *judging*, *suspecting*, *assessing*, *realizing*, *understanding*, *believing* or *desiring*, etc., that *p*.<sup>5</sup> Thus, any subsequent mention of thinking will refer to the *determinable* of which judging, suspecting, assessing, realizing, understanding, believing, desiring, etc., are *determinates*. Naturally, all the determinates of which thinking is a determinable can be non-accidentally *about oneself*. Consequently, in my terminology, thinking about oneself *as oneself* broadly encompasses judging, suspecting, assessing, realizing, understanding, believing, desiring, etc., *that p* – where the proposition *p* involves a reference to oneself *as oneself*. With these preliminary distinctions in place, I will now move to the question of whether engaging in conscious *de se* thinking qualifies as a form of phenomenal self-consciousness.

## 5.2 *De se* thought and self-consciousness

Recall that in chapter 4 I defined *phenomenal self-consciousness* as an egological notion of self-consciousness (consciousness *of oneself*) stronger than Legrand's and Nida-Rümelin's minimal notion of 'pre-reflective self-consciousness'. We can now ask the following question: is engaging in *de se* thinking sufficient to meet the definitional requirements for

<sup>5</sup>Not everyone would agree that non-conscious *belief*, for example, is an instance of non-conscious *thinking*, if non-conscious belief is treated as a dispositional state, and thinking is treated as an episode that unfolds over time (see for example Crane 2013). We need not get into this debate here.

phenomenal self-consciousness? If it is possible at all to engage in *non-conscious de se* thinking (for example, by thinking non-consciously that one *oneself* is underdressed), then the answer to this question is presumably negative. Indeed, it follows from (PSC) that a subject *S* who is phenomenally self-conscious at *t* is *ipso facto* in a *conscious* mental state at *t*, that is, in a state of consciousness of *oneself as oneself* that constitutively contributes to *S*'s overall phenomenology. Consequently, being in a mental state *M* cannot meet the definitional requirement for phenomenal self-consciousness if *M* is not a *conscious* mental state. It follows that in order for a subject engaging in *de se* thinking to be (*ipso facto*) phenomenally self-conscious, it must be the case that the subject is engaging in *conscious* – rather than *non-conscious* – *de se* thinking.

Until recently, it was commonly assumed by philosophers in the analytic tradition that while there is something it is like to have perceptual experiences, feelings of pain and pleasure, or emotions, there is nothing it is like to engage in thinking. Let us call this general claim the conservative view of the phenomenology of thinking. On a strong version of the conservative view, engaging in thinking never makes any contribution to the subject's phenomenology; consequently, thinking is never a conscious mental state. Here is a representative example of the strong conservative view:<sup>6</sup>

Bodily sensations and perceptual experiences are prime examples of states for which there is something it is like to be in them. They have a phenomenal feel, a phenomenology, or, in a term sometimes used in psychology, raw feels. Cognitive states are prime examples of states for which there is not something it is like to be in them, of states that lack a phenomenology.

Braddon-Mitchell and F. Jackson (2007, p. 129)

On a seemingly weaker – and more widespread – version of the conservative view, engaging in thinking does not involve any *distinctive* or *proprietary* phenomenology, but it may nonetheless make a contribution to the subject's phenomenology in so far as it may bring about perceptual, algedonic or emotional phenomenology. Here is a representative example of this view:<sup>7</sup>

<sup>6</sup>See also Nelkin (1989, p. 430).

<sup>7</sup>See also Carruthers and Veillet (2011) and Tye and Wright (2011).

Our thoughts aren't like anything, in the relevant sense, except to the extent that they might be associated with visual or other images or emotional feelings, which will be phenomenally conscious by virtue of their quasi-sensory status.

Carruthers (2005, pp. 138-9)

Note that the difference between these two versions of the conservative view is not as clear-cut as it might seem at a first pass. Indeed, both positions appear to state that engaging in thinking does not *constitutively* contribute to the subject's overall phenomenology. The weaker conservative view simply seems to emphasise that engaging in thinking may *indirectly* contribute to the subject's overall phenomenology, to the extent that it may causally bring about distinct mental states, such as sensory or emotional experiences, that do make a constitutive contribution to phenomenology. However, this claim is not necessarily incompatible with the 'strong' version of the conservative view, which does not deny that sensory and emotional states constitutively contribute to the subject's overall phenomenology – regardless of how they are formed. This distinction is familiar from chapter 1: if thinking about my dog Fido systematically causes me to feel joy, then thinking about Fido *causally* – but not *constitutively* – contributes to my overall phenomenology. In such a case, the change in my overall phenomenology brought about by my thinking about Fido would be exhausted by the emotional phenomenology of the joy caused by my thinking about Fido. Thus, both versions of the conservative view can agree that my thinking about Fido does not *constitutively* contribute to my overall phenomenology. Generally, both versions can agree that a subject S's thinking *that p* is always a *non-conscious* mental state of S, in so far as it does not *constitutively* contribute to S's overall phenomenology.

Nonetheless, it is possible to articulate a version of the conservative view that is clearly distinct from the stronger variant, by saying that one can engage in *conscious* thinking, where this means that one's thinking *that p* can make a *constitutive* contribution to one's overall phenomenology even though it does not have *proprietary* phenomenology. For example, one might say that some (or all) instances of thinking have 'quasi-auditory' phenomenology in so far as they involve inner speech (Prinz 2011).<sup>8</sup> The idea is not that engaging

<sup>8</sup>By 'quasi-auditory phenomenology', I refer to the phenomenology of inner speech in so far as what it is like to 'hear' a sentence *in one's head* differs from what it is like to hear the same sentence as spoken by someone in one's environment.

in thinking may causally bring about *distinct* mental states with sensory phenomenology, but that the very state of thinking can constitutively contribute sensory phenomenology to the subject's overall phenomenology. On this view, there is constitutively something it is like to think *that p*, in so far as thinking *that p* is not merely causally "associated with visual or other images or emotional feelings" (Carruthers 2005, pp. 138-9); however, what it is like to think *that p* does not pertain to a proprietary phenomenology of thinking, but to the quasi-auditory phenomenology of inner speech.

There is some daylight between the view that thinking may constitutively involve *non-proprietary* phenomenology (e.g., quasi-auditory phenomenology), and the view that thinking may *merely causally* bring about sensory states but does not make a constitutive contribution to phenomenology. For the sake of clarity, we might call the latter the 'intermediate' conservative view, and the former the 'weak' conservative view properly speaking (see figure 5.3 for a summary). In practice, the distinction between the 'intermediate' and the 'weak' versions of the conservative view might come down to considerations about the individuation of mental states. Suppose, for example, that I think that Paris is the capital of France, and that upon having that thought I experience a quasi-auditory phenomenology of inner speech corresponding to the sentence "Paris is the capital of France". Some might consider that this scenario involves two distinct mental states – a *non-conscious* cognitive state without phenomenology and a *conscious* sensory state with phenomenology –, while others might consider that this scenario involves a single mental state – a *conscious* cognitive state with sensory phenomenology. Nonetheless, the 'intermediate' and 'weak' versions of the conservative view are distinct in principle.

The various versions of the conservative view have recently come under criticism from proponents of the liberal view, according to which engaging in thinking does involve a proprietary phenomenology, typically called 'cognitive phenomenology' (see figure 5.3). Here is a representative example of the liberal view:<sup>9</sup>

In addition to arguing that there is something it is like to think a conscious thought, I shall also argue that what it is like to think a conscious thought is distinct from what it is like to be in any other kind of conscious mental state, and that what it is like to

<sup>9</sup>See also Strawson (1994), Siewert (1998), T. E. Horgan et al. (2004), and Chudnoff (2015a).

think the conscious thought that *p* is distinct from what it is like to think any other conscious thought...

Pitt (2004, p. 2)

I will not get into the intricacies of the debate between these different positions here; rather, I will briefly explore the implications of each position for the relationship between *de se* thinking and phenomenal self-consciousness.

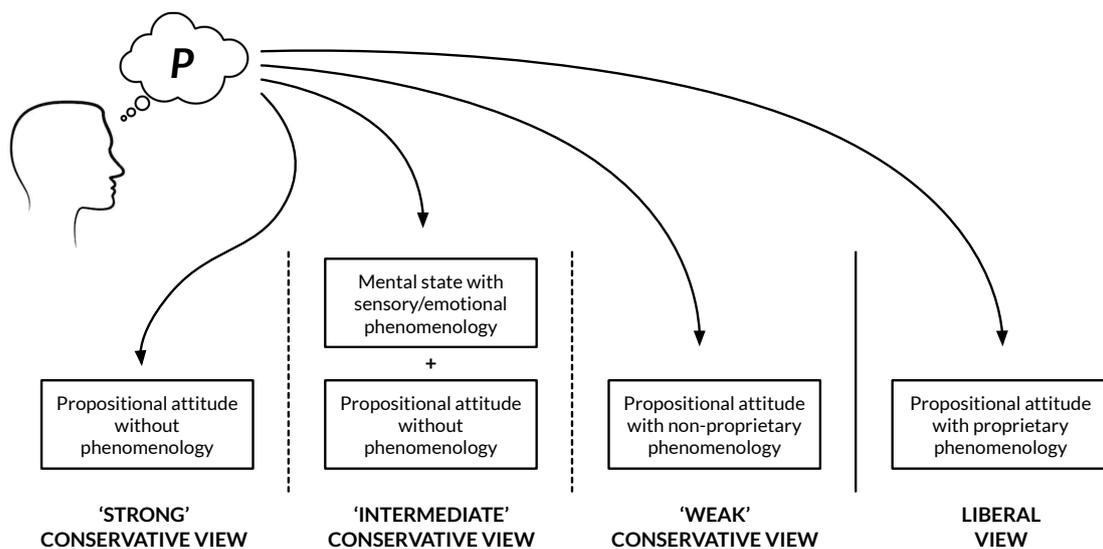


Figure 5.3: Different views regarding the phenomenology of thinking.

Recall that for a subject *S* to be phenomenally self-conscious, as defined by (PSC), *S* must be conscious of oneself (as oneself), and *S*'s being conscious of oneself (as oneself) must make a specific constitutive contribution to *S*'s overall phenomenology. Consequently, the issue at hand is whether subject *S*'s thinking of oneself as oneself meets the definitional requirements for being conscious of oneself as oneself in such a way that *S*'s being conscious of oneself as oneself makes a constitutive contribution to *S*'s overall phenomenology. In other words: if *S* is engaged in *de se* thinking, is *S ipso facto* phenomenally self-conscious?

Being phenomenally self-conscious is a conscious mental state but, according to the 'strong' and 'intermediate' versions of the conservative view, subject *S*'s engaging in *de se* thinking makes no constitutive contribution to *S*'s overall phenomenology. Thus, it is

immediate that, on these versions of the view, a subject engaged in *de se* thinking is not *ipso facto* phenomenally self-conscious. On the ‘weak’ conservative view, subject S’s engaging in *de se* thinking can make a constitutive contribution to S’s overall phenomenology; and what it contributes is the phenomenology of inner speech. This need not always be the case when S is engaging in *de se* thinking, as some proponents of the ‘weak’ conservative might want to maintain that S may *non-consciously* think *that p*. For example, in some particular circumstance, S might *non-consciously* think that S [SELF] is underdressed. Nonetheless, the ‘weak’ conservative view allows that some instances of *de se* thinking are *conscious*, in so far as they constitutively contribute the phenomenology of inner speech to their subject’s overall phenomenology. Thus, on the ‘weak’ conservative view, a subject engaged in *conscious de se* thinking is *ipso facto* phenomenally self-conscious only if the phenomenology of inner speech is or includes the phenomenology of self-consciousness. There are reasons to doubt that this is the case.

Consider the difference between consciously thinking “I am making a mess” and consciously reading the sentence “I am making a mess”. According to the ‘weak’ conservative view, both experiences might involve the phenomenology of inner speech, that is, they might involve quasi-hearing the sentence ‘I am making a mess’ in one’s head. Furthermore, the experience of consciously thinking “I am making a mess” should not make any contribution to one’s overall phenomenology that the experience of consciously reading “I am making a mess” does not also make, since the former simply involves the phenomenology of inner speech. For a subject S to be phenomenally self-conscious, in the sense defined by (PSC), it is presumably not sufficient for S to quasi-hear in her head a sentence that happens to include the first-person pronoun. Indeed, one may read or hear the sentence “I am making mess” without consciously representing what this sentence means. More specifically, the experience of hearing the phoneme [ aɪ ] (‘I’) in one’s head is not sufficient to make one conscious of oneself *as oneself*, unless one also entertains the meaning of this phoneme in English in such a way that (a) one represents oneself *as oneself* by making use of the first-person concept, and (b) the deployment of the first-person concept contributes to one’s overall phenomenology.

Admittedly, it does presumably follow from the ‘weak’ conservative view – and, for that matter, from all versions of the conservative view – that engaging in *de se* thinking involves representing oneself *as oneself* by making use of the first-person concept. However, such self-representation should not be consciously registered by the subject if the only contribution of *de se* thinking to the subject’s overall phenomenology is a matter of quasi-auditory inner speech. Thus, it is not exaggerated to say that on the ‘weak’ conservative view, conscious thinking is “just and wholly a matter of sensation... on the experience side, and entirely non-conscious registration of meaning... on the non-experiential side.” (Strawson 2011a, p. 295). Specifically, conscious *de se* thinking is just a matter of quasi-auditory sensation within the subject’s overall phenomenology, and a matter of non-conscious self-representation outside of the subject’s overall phenomenology. In other words, the various versions of the conservative view entail that *de se* thinking involves *self-representation*, but not quite *self-consciousness*.

On the liberal view, a subject S’s engaging in thinking can also make a constitutive contribution to S’s overall phenomenology; but what it contributes is the proprietary phenomenology of thinking (cognitive phenomenology), rather than the phenomenology of inner speech. Cognitive phenomenology is intimately related to the subject’s grasp of the thought’s propositional content. Indeed, as Montague (2016a) puts it, it is reasonable to assume that “the phenomenology that makes a particular occurrent thought a conscious thought must be explanatorily or intelligibly linked to the representational content of that thought” (p. 171).

As we have seen, it is difficult to link the quasi-sensory phenomenology that may accompany a thought (e.g. through inner speech) to the representational content of that thought: the phenomenology of quasi-hearing the phoneme [ ai ] does not seem intelligibly related to the representation of oneself as oneself through the deployment of the first-person concept. The liberal view addresses this by postulating the existence of non-sensory cognitive phenomenology, where such phenomenology is tightly connected to the content of thoughts.<sup>10</sup> While the representation of oneself as oneself in *de se* thinking is a common

<sup>10</sup>For the *representationalist* version of the liberal view, the (proprietary) phenomenology of a thought is fully determined by the thought’s intentional content (Bourget 2017). For the *phenomenal intentionalist* ver-

factor across all the views of the phenomenology of thinking, what matters for phenomenal self-consciousness is that *de se* thinking should involve *conscious* self-representation, namely a *phenomenology* of representing oneself as oneself. The liberal view satisfies this requirement: what it is like to consciously think a *de se* thought, on this view, includes the cognitive phenomenology of representing oneself as oneself. Consequently, according to the proponent of the liberal view, not only does a subject S's engaging in conscious *de se* thinking make a constitutive contribution to S's overall phenomenology (as is also the case on the 'weak' conservative view), but by being engaged in conscious *de se* thinking S is *ipso facto* phenomenally self-conscious.

I have argued so far that consciously thinking about oneself as oneself can only count as an instance of phenomenal self-consciousness in so far as thoughts have a cognitive phenomenology that is intelligibly related to their contents (as the liberal view suggests). While I cannot offer a detailed defence of the liberal view here, I shall briefly explain why I think it is more plausible than alternatives. First, it is helpful to say a few words about how might one positively characterise the cognitive phenomenology of thinking. I have already noted that the cognitive phenomenology postulated by proponents of the liberal view is by definition irreducible to sensory phenomenology; it is not a matter of 'quasi-hearing' words or 'quasi-seeing' images in one's head as one thinks. Rather, one might say that it is the experience of *meaning*, or perhaps more aptly the experience of entertaining and understanding – holding before one's mind, as it were – the propositional content of one's conscious thoughts.

Suppose that you are Perry (1979)'s messy shopper, pushing a shopping cart containing a leaking bag of sugar down the aisle of a grocery store. As you go around the store, you start noticing that there is a trail of sugar on the floor, and you think "Someone is making a mess". Every time you go around, the trail becomes thicker, and you cannot seem to find who the messy shopper is. After racking your brain to figure out who in the store is the messy shopper, you have a sudden realization: you think "*I* am making a mess".

sion of the liberal view, the intentional content of a thought is fully determined by the thought's (proprietary) phenomenology (Pitt 2004). I will not get into the debate between representationalism and the phenomenal intentionality view here.

Presumably, this is a conscious and salient thought – it is the culmination of a train of voluntary and effortful thoughts attempting to solve the mystery of the messy shopper. It is intuitively plausible that having the realization that you are the messy shopper would make a constitutive contribution to your overall phenomenology; this contribution is additional to whatever else you might experience at the time, for example your visual experience of the trail of sugar on the floor. Furthermore, there seems to be more to such an experience than hearing (or quasi-hearing) in your head the words “I am making a mess” in inner speech. The auditory imagery of the phonemes that make up these words does not, by itself, *mean* anything. By contrast, having the realization that you are the messy shopper is a meaningful conscious episode. According to the liberal view, what it is like to think “I am making a mess” involves cognitive phenomenology, namely the phenomenology of entertaining and understanding the meaning of the proposition “I am making a mess”.

There are three main classes of arguments in favour of the liberal view. A first class of arguments appeals to examples of familiar experiences whose phenomenology is difficult to account for in merely sensory or emotional terms. Such examples include suddenly remembering something (e.g., suddenly remembering where one has put one’s car keys), tip-of-the-tongue phenomena, and being immersed in a stream of thoughts. My use of the example of the messy shopper (which diverges from Perry’s use of the example) would also fit within this first category. Indeed, it appeals to the presumably familiar experience of having a sudden realisation after pondering a problem for a while. This is a mental event that is phenomenologically salient; and its phenomenology does not seem reducible to hearing (or quasi-hearing) words in one’s mind.

A second class of arguments in favour of the liberal view appeals to the phenomenal contrast between minimal pairs of experiences, namely pairs of experiences whose respective phenomenal characters seem to differ only in that one of them involves cognitive phenomenology while the other lacks such phenomenology. Examples of relevant phenomenal contrasts includes the contrast in one’s experience before and after one grasps a mathematical proof, or the contrast between hearing a sentence in a foreign language without understanding it and hearing the same sentence when one understands it. Thus, Strawson (1994) considers the respective experiences of Jack (a monolingual English speaker) and

Jacques (a monolingual French speaker) as they listen to the news in French. Suppose Jack and Jacques both hear the sentence “Le chef d’État s’est rendu à l’étranger” on the radio. Jack’s and Jacques’ respective experiences both involve the same auditory phenomenology (assuming their hearing is equally good); however, Jacques’ experience will presumably involve some additional phenomenology, namely the phenomenology of understanding the meaning of the sentence “Le chef d’État s’est rendu à l’étranger” (which is translated as “The Head of State has travelled abroad” in English). This phenomenology of understanding is what the cognitive phenomenology of thinking is taken to be: when one consciously thinks *that the Head of State has travelled abroad*, one does not – or not merely – hear the words that compose the corresponding English sentence in one’s mind; there is (also) something it is like to entertain and understand the meaning of the proposition *the Head of State has travelled abroad*.

A third class of arguments in favour of the liberal view relies on *a priori* considerations rather than introspection or phenomenal contrast. Thus, Montague (2016a) plausibly suggests that for a subject S’s thinking *that p* to be a *conscious* mental state of S – that is, for S to engage in *conscious* thinking *that p* –, S must consciously entertain the propositional content *p* of that thought. From there, she argues that an adequate account of what it is for the propositional content of a thought to be consciously entertained must appeal to cognitive phenomenology, because merely sensory or quasi-sensory phenomenology (e.g., the phenomenology of inner speech) cannot represent the propositional content of thoughts. Consequently, cognitive phenomenology – the phenomenology of entertaining the meaning of propositions – must be constitutive of conscious thinking, if engaging in thinking is ever to make a constitutive contribution to a subject’s overall phenomenology. I find this argument plausible: it is antecedently plausible that thinking *that p* can constitutively contribute to one’s overall phenomenology, and that what thinking *that p* contributes to one’s overall phenomenology must involve what it is like to entertain of the propositional content *p*; since it is also plausible that sensory or quasi-sensory phenomenology alone cannot account for what it is like to entertain of the propositional content *p*, consciously thinking *that p* must involve cognitive phenomenology (the phenomenology of holding the meaning of *p* before one’s mind).

Once we accept the liberal view, we can formulate the following argument in favour of the claim that any subject S engaged in conscious *de se* thinking is *ipso facto* phenomenally self-conscious:<sup>11</sup>

- (P1) A subject engaged in conscious *de se* thinking is *ipso facto* phenomenally self-conscious if and only if engaging in conscious *de se* thinking involves consciously representing oneself as oneself.
- (P2) If a subject consciously entertains the meaning of a proposition featuring the concept of self, then she consciously represents herself as herself.
- (P3) A subject engaged in conscious thinking *that p* consciously entertains the meaning of *p* if and only if conscious thinking *that p* has cognitive phenomenology.
- (P4) All conscious thinking has cognitive phenomenology.
- (C1) Therefore, a subject engaged in conscious thinking *that p* consciously entertains the meaning of *p*.
- (P5) Conscious *de se* thinking is conscious thinking *that p*, where *p* features the concept of self.
- (C2) Therefore, a subject engaged in conscious *de se* thinking consciously entertains the meaning of a proposition featuring the concept of self.
- (C3) Therefore, a subject engaged in conscious *de se* thinking consciously represents herself as herself.
- (C4) Therefore, A subject engaged in conscious *de se* thinking is *ipso facto* phenomenally self-conscious.

<sup>11</sup>Note that one could avoid the reference to cognitive phenomenology altogether by endorsing (C1), the claim that a subject engaging in conscious thinking *that p* is *ipso facto* consciously entertaining the meaning of *p*, as a premise (replacing (P3)). While I do find (C1) antecedently plausibly, I think the comparison of the 'weak' conservative view with the liberal view is helpful to get a better grip on what it means for a subject to consciously entertain the meaning of a proposition.

(P1) plausibly follows from the definition of phenomenal self-consciousness (PSC). (P2) is antecedently plausible if one agrees that consciously entertaining the meaning of a proposition involves consciously entertaining the meaning of the concepts featured in that proposition, and that consciously entertaining the meaning of the concept of self involves consciously representing oneself as oneself. (P3) is plausible in light of the consideration offered by Montague (2016a) in favour of the liberal view: given that cognitive phenomenology is the phenomenology of entertaining the meaning of a proposition, a subject engaged in conscious thinking *that p* cannot consciously entertain the meaning of *p* if her conscious thinking *that p* does not have cognitive phenomenology. (P4) simply is the liberal view of the phenomenology of thinking that has already been defended. Finally, (P5) straightforwardly follows from the definition of *de se* thinking.

### 5.3 Is cognitive self-consciousness ubiquitous?

I have argued so far that when one consciously thinks of oneself as oneself, one is *ipso facto* phenomenally self-conscious. In chapter 4, I suggested that there are no compelling arguments for the claim that consciousness *constitutively* involves phenomenal self-consciousness. If this is right, then there should be no compelling argument for the more specific claim that consciousness *constitutively* involves conscious *de se* thinking. To my knowledge, no scientific or philosophical theory of consciousness holds that being conscious at all necessarily involves consciously thinking, let alone consciously thinking about oneself as oneself.<sup>12</sup> One can occasionally find formulations that suggest that being

<sup>12</sup>Admittedly, the so-called Higher-Order Thought (HOT) theory of consciousness does claim that conscious mental states are systematically represented by higher-order thoughts, such that being conscious constitutively involves having thoughts about one's first-order mental states (Rosenthal 1986; 1993; 2005). Furthermore, some versions of the HOT theory hold that the relevant higher-order thoughts are in fact *de se* thoughts: "[the HOT's] content must be that one is, *oneself*, in [the first-order] mental state... HOTs refer both *to oneself* and to one's mental states" (Rosenthal 1997, p. 741, my emphasis). However, the HOT theory is not committed to the claim that higher-order thoughts about first-order mental states should themselves be *conscious* thoughts; its proponents readily acknowledge that "the HOTs are typically not conscious thoughts" (Rosenthal 1997, p. 745). Thus, the HOT theory of consciousness does not vindicate the claim that consciousness constitutively involves conscious *de se* thought, nor does it vindicate the broader claim that consciousness constitutively involves phenomenal self-consciousness.

conscious constitutively involves engaging in conscious thoughts in general. Perhaps the following passages from O'Shaughnessy (2000) offer an example of such a view:

The mind of one who is conscious is *necessarily* a mind actively governing the movement of its own attention and *thinking processes*.

O'Shaughnessy (2000, p. 89, my emphasis)

In the final analysis it is because thinking is active and *thinking is essential to consciousness* that mental action is a necessary condition of consciousness... Thus thinking [is] the source of the interiority necessary to consciousness...

O'Shaughnessy (2000, p. 264, my emphasis)

However, in the context in which these passages occur, O'Shaughnessy appears to restrict this claim to the kind of consciousness one has in the ordinary wakeful state. More generally, there is no reason to believe that in order to be conscious at all, one should engage in conscious thinking, unless one simply equates 'thought' with 'experience' as early modern philosophers often did. *A fortiori*, there is no reason to believe that in order to be conscious at all, one should engage in conscious *de se* thinking. While it is very plausible that a number of non-human animal species have the capacity to be conscious, it is much less plausible that all members of the relevant species have the capacity to think of themselves as themselves. Furthermore, even if we focus on human subjects, it would be odd to claim that all such subjects are perpetually thinking of themselves as themselves whenever they are conscious, from the moment they wake up in the morning to the moment they slip into unconsciousness at night. Even the most narcissistic individuals presumably live through many conscious episodes in which they are not thinking about themselves in any way.

While these remarks are intuitively obvious, they can be substantiated by empirical evidence regarding states of consciousness lacking conscious *de se* thinking. I will focus here on relevant evidence concerning three conditions: so-called flow states, states induced by meditation practice, and dreaming. It has been argued that the performance of attention-demanding activities can induce a state of *flow*, especially at expert level, characterized by extreme and undivided focus on the task at hand (Csikszentmihalyi 1990). The Flow State Scale (FSS) was designed to measure different aspects of flow states across nine factors (S. A. Jackson and Marsh 1996). In a sample of 747 respondents from a variety of competitive and noncompetitive sport activities, the mean score (on a five point Likert scale from 1 =

‘strongly disagree’ to 5 = ‘strongly agree’) for the FSS factors *Concentration on the task at hand* (measuring the subject’s focus and immersion in her activity) and *Loss of self-consciousness* (measuring the “absence of preoccupation with self” and the degree to which the subject “is not focusing on the information normally used to represent to oneself who one is”<sup>13</sup>) were 3.73 and 3.92 respectively (S. Jackson and Eklund 2004). Similar results were obtained with a sample of 236 music student performing for an examination, with mean scores of 3.53 and 3.12 for the same FSS factors (Wrigley and Emmerson 2013). These data suggest that skilled individuals focused on a demanding performance may reach a state in which they are entirely focused on task-related goals, and therefore are unlikely to engage in conscious *de se* thinking, let alone *constantly* engage in such thinking.

Dreaming is another condition in which cognitive self-consciousness might temporarily subside. An analysis of 500 rapid eye movement (REM) dreams in the laboratory found that only 21.2% of dream reports included instances of thinking, and that such instances only made up 5% of the content of reported dreams (Meier 1993). A more recent study obtained reports about 788 dreams from 144 participants over two years using the Subjective Experiences Rating Scale (SERS). On average, participants gave a score of 2.29 for the item “During how much of the time were you engaged in thinking?” on a scale from 0 (“none of the time”) to 4 (“all of the time”) (Kahan and Claudatos 2016) – indicating that a significant proportion of most dreams lack thoughts. These results suggest that “different forms of thinking can be lacking in dreams” (Windt 2015, p. 424), including thinking of oneself as oneself.

While questionnaire reports about flow states and dreams should be interpreted with caution, research on meditation provides stronger evidence that cognitive self-consciousness is not ubiquitous in all conscious experiences. Many meditation practices, especially in the Buddhist and Hindu traditions, explicitly aim at suppressing spontaneous thoughts and focusing attention on sensory stimuli.<sup>14</sup> This is the case in particular for

<sup>13</sup>S. A. Jackson and Marsh (1996, p. 19).

<sup>14</sup>“[T]he original goal of Meditation is the elimination or reduction of thought processes, the cessation or slowing of the internal dialogue of the mind” (Rubia 2009, p. 2); “Meditation... is aimed to reduce thoughts to ultimately reach the state of thoughtless awareness which is considered a different state of consciousness where one is fully perceptually alert, yet has no thoughts. It is consequently described as a state of pure

a large family of meditation practices commonly called ‘Focused Attention meditation’. Focused Attention meditation involves sustaining one’s attentional focus on a particular object, either internal (e.g., breathing sensations) or external (e.g., ambient sounds or the flame of a candle). The practitioner is instructed to monitor their attention, notice episodes of distraction (mind-wandering), and subsequently bring their attention back to the object of the meditation. More specifically, the practice of Focused Attention meditation cycles indefinitely through four phases: (1) sustained attention on a particular object or sensory stimulus (typically the breath), (2) occurrence of spontaneous task-unrelated thoughts (mind-wandering), (3) awareness of the mind-wandering episode (also called ‘meta-awareness’), (4) re-focusing of attention on the relevant object or sensory stimuli (see figure 5.4).

It is very plausible that the second phase of the practice, during which meditators get distracted and let their minds drift into spontaneous thoughts, routinely involve conscious *de se* thinking – for example, thinking that one is hungry. The third phase, in which meditators ‘catch’ themselves in the act of mind-wandering, also plausibly involves conscious *de se* thinking, such as the thought “I am lost in thoughts again”. It is less clear whether the fourth phase involves conscious *de se* thinking, since it corresponds to the transition period during which meditators bring their attention back to the object or stimulus on which it should be focused. In any case, the first phase of the practice, during which the meditator’s attention is fully focused, is intended to be free of spontaneous thoughts, including *de se* thoughts (Hasenkamp 2018). Indeed, advanced meditators report that once attention has been stabilized and the mind ‘quieted’, they can momentarily undergo conscious episodes lacking conscious thoughts altogether.<sup>15</sup>

attention without any thought content.” (Hernández et al. 2016, p. 2); “[T]he Hindus had postulated, defined and demonstrated the existence of a fourth... state of consciousness which is neither ordinary wakefulness, dream nor sleep. What is this fourth... state? This has been defined as a state where the mind is still and there are no thoughts.” (Ramamurthi 1995, p. 108).

<sup>15</sup>“My mind became more free and I became more and more peaceful, without any thoughts” (Siff 2014); “You’re aware of your whole body... at once, without any thoughts at all” (Selby 2003); “[T]houghts seemed to become confused as they faded into nothingness. I experienced myself as being in a thoughtless and timeless place” (Barnes 2001, p. 7).

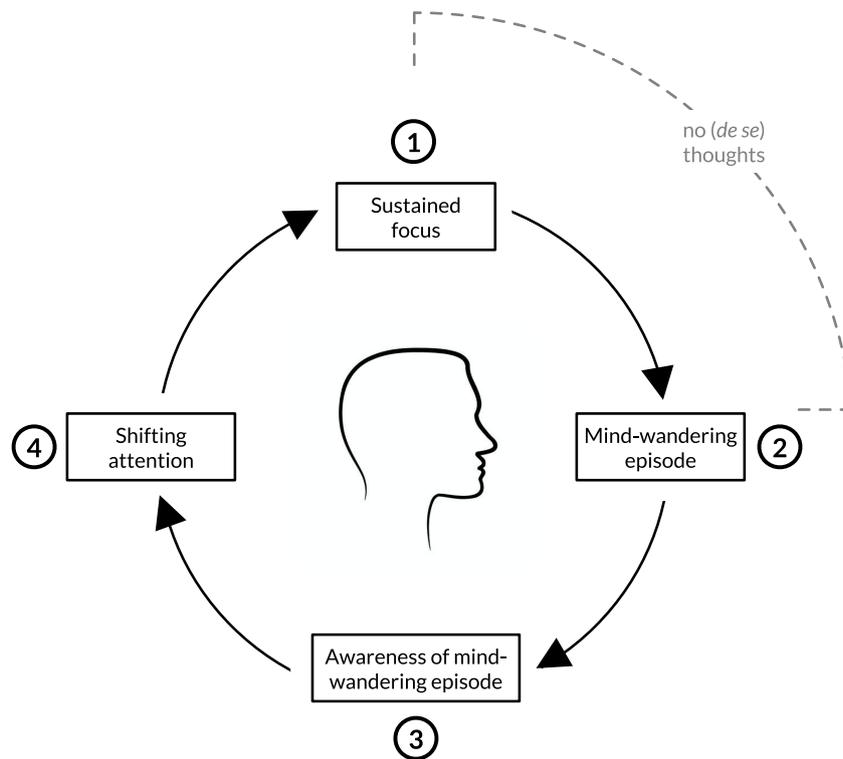


Figure 5.4: The phases of Focused Attention meditation (adapted from Hasenkamp et al. 2012).

These remarks should suffice to show that being conscious at all does *not* constitutively involve engaging in conscious *de se* thinking; furthermore, conscious *de se* thinking is not even ubiquitous in all actual conscious experiences as a matter of contingent empirical fact. I will now assess how prevalent cognitive self-consciousness actually is in ordinary conscious experience.

#### 5.4 Cognitive self-consciousness in ordinary experience

After arguing against (CSC<sub>E</sub>) in chapter 4, I endeavoured to shed light on the commonly held view that phenomenal self-consciousness, although not *constitutive* of consciousness, might nonetheless be relatively pervasive in conscious experience, particularly in a set of ‘ordinary circumstances’ to be further characterised:

(TSC) *In ordinary circumstances*, if a subject S is conscious, (a) S is conscious of S [SELF], and (b) S’s being conscious of S [SELF] makes a determinate constitutive contribution to S’s overall phenomenology at *t*.

I have suggested that the difficulty of assessing this claim lies partly in the indeterminacy of the relevant notion of phenomenal self-consciousness. We now know that conscious *de se* thinking can plausibly be taken to be a determinate of which phenomenal self-consciousness is a determinable, and we can ask how prevalent conscious *de se* thinking is in conscious experience. Of course, unless conscious *de se* thinking is the *only* way in which one can be phenomenally self-conscious, answering that question will not be sufficient to assess (TSC). Nonetheless, conscious *de se* thinking is *one* way in which one can be phenomenally self-conscious, and consequently assessing its prevalence in conscious experience should at least bring us one step closer to assessing (TSC).

While it is difficult to precisely assess the prevalence of conscious *de se* thinking in ordinary experience, the empirical literature on mind-wandering provides some insight into this question. Mind-wandering refers to the spontaneous occurrence of thoughts in the waking state, and particularly thoughts that are not related to the specific task that one is currently performing (if any). There is converging evidence that episodes of mind-wandering are more frequent in ordinary experience than we might think. Studies using experience sampling<sup>16</sup> found that mind-wandering may occur in more than 40% of the samples obtained by probing the mental life of participants at random intervals (Christoff et al. 2009; Killingsworth and Gilbert 2010). Furthermore, we often fail to notice when our minds are wandering, which might cause us to underestimate how often we are lost in thoughts (Schooler et al. 2011).

What is the proportion of *de se* thought during mind-wandering? Many experimental studies have indicated that the content of task-unrelated thoughts is often related to oneself (Andrews-Hanna et al. 2013; Baird et al. 2011; Ruby, Smallwood, Engen, et al. 2013; Ruby, Smallwood, Sackur, et al. 2013; Song and Wang 2012; Stawarczyk, Cassol, et al. 2013). There is also considerable evidence that mind-wandering episodes are predominantly focused on the future, and frequently involve the planning and anticipation of personal goals, known as autobiographical planning (Baird et al. 2011; D'Argembeau

<sup>16</sup>Experience sampling involves probing participants at random intervals and asking them to report on aspects of their subjective experience immediately before the probe (Hurlburt and Akhter 2006).

2016; Ruby, Smallwood, Engen, et al. 2013; Ruby, Smallwood, Sackur, et al. 2013; Smallwood, Nind, et al. 2009; Smallwood, Schooler, et al. 2011; Stawarczyk, Cassol, et al. 2013; Stawarczyk, Majerus, et al. 2011). Furthermore, task-unrelated thoughts focused on the future seem particularly likely to be related to the self.

For example, Baird et al. (2011) assessed the proportion of self-related thoughts in participants who were pseudo-randomly prompted to report their thoughts while performing a task. They found that 77% of task-unrelated thoughts that were focused on the past and as much as 92% of task-unrelated thoughts that were focused on the future were related to the self. As the authors indicate, “thought samples were classified as self-related [by four independent judges] if they included specific mention of an individual’s self” (p. 1606) – presumably through the use of the first-person pronoun in the participants’ reports. It is plausible that most if not all reports classified as ‘self-related’ in that sense did correspond to conscious thoughts about oneself as oneself (e.g., “I need to buy milk for breakfast”, or “I should apply for this job”).

Thus, there is convincing evidence that conscious *de se* thoughts might be more prevalent in ordinary experience than we notice: we spend a significant part of our lives lost in thoughts about ourselves, particularly when we spontaneously think about future actions and goals (see D’Argembeau 2018 for a discussion). Nonetheless, this point should not be overstated, for experience sampling studies also suggest that a large portion of our mental lives is *not* spent lost in task-unrelated thoughts. Of course, thoughts still occur outside of mind-wandering episodes, in which case they are generally related to the specific task that subjects are performing (e.g., finding one’s way in the subway, solving a puzzle, or writing a doctoral thesis). However, it is plausible that many task-related thoughts are not focused on oneself.<sup>17</sup>

In their experiment, Baird et al. (2011) found that only 46% of all *on task* (task-related) thought samples were self-related, compared to 66% of all *off task* (task-unrelated) thought

<sup>17</sup>Try solving the following equation:

$$x^2 + 5 = 13 - 4$$

Once you have determined the value of  $x$ , ask yourself whether you were thinking about yourself in any way while you were pondering the problem. I predict that the answer will be almost unanimously negative.

samples. Although they do not indicate the proportion of all thought samples that were on task and off task respectively, we know from aforementioned experience sampling studies that 40% to 50% of our waking lives is spent mind-wandering (Killingsworth and Gilbert 2010). Thus, if we assume that roughly 50% of thought samples from Baird et al. (2011)'s experiment were off task, this means that about 56% of *all* thought samples they collected were self-related. In so far as we can further extrapolate from these data, we can tentatively speculate that roughly *half* of all our conscious thoughts are *de se* thoughts. Thus, given that we are probably not constantly engaging in conscious thoughts from the moment we wake up to the moment we go to sleep, it very plausible that we spend *less than half* of our waking lives – and perhaps significantly less than half of our waking lives – thinking about ourselves as ourselves.

Coming back to (TSC), this analysis suggests that even if we interpret the relevant set of 'ordinary circumstances' in a very restrictive manner – for example restricting the claim to the ordinary wakeful state of consciousness of healthy neurotypical adult human beings – the claim would be empirically wrong if the only way in which one could be phenomenally self-conscious was by consciously thinking of oneself as oneself. Not only is cognitive self-consciousness not constitutively involved in all consciousness, but it is also far from ubiquitous even in the ordinary conscious experience of normal adult human beings. This raises the obvious question of whether cognitive self-consciousness is the *only* way in which one can be phenomenally self-conscious.

## 5.5 Conclusion

At the outset of this chapter, I mentioned that a number of authors seem to assimilate the ability to be self-conscious to the capacity for *de se* thoughts. I have subsequently argued that if conscious *de se* thinking is indeed the only form of phenomenal self-consciousness, then (TSC) appears to be empirically inaccurate, even on a restrictive understanding of the relevant set of 'ordinary circumstances'.

In the following chapters, I will offer new arguments in favour of the claim that cognitive self-consciousness is *not* the only form of phenomenal self-consciousness. Indeed, I will argue that there are at least two other determinate notions of which phenomenal

self-consciousness is a determinable: *bodily self-consciousness*, namely the consciousness of oneself as the bodily subject whose body part one feels a sensation in or on, and *spatial self-consciousness*, namely the consciousness of oneself as being located at the origin of the spatial perspective one's perceptual experience. I will ask two questions regarding each of these notions. Firstly, I will consider whether each notion refers to an actual feature of experience whose phenomenology is distinct from that of cognitive self-consciousness. Secondly, after answering the first question affirmatively, I will ask what is the actual prevalence of bodily self-consciousness and spatial self-consciousness in conscious experience. This will allow me to assess (TSC) again, and defend the empirical accuracy of the claim, at least on restrictive understanding of the relevant set of 'ordinary circumstances'.

## 6

# Bodily Self-Consciousness

Like the shadow of one's own hand, [the self] will not wait to be jumped on... It evades capture by lodging itself inside the very muscles of the pursuer.

Ryle (1949, p. 186)

In the previous chapter, I argued that a subject engaged in conscious *de se* thinking is *ipso facto* phenomenally self-conscious. I have called the determinate kind of phenomenal self-consciousness that consists in engaging in conscious *de se* thinking *cognitive self-consciousness*. I suggested that cognitive self-consciousness is not ubiquitous in ordinary experience: we frequently undergo conscious episodes during which we do not consciously think about ourselves as ourselves. If being cognitively self-conscious is the *only* way in which we can be phenomenally self-conscious, then we should reject (TSC), the typicalist claim that that phenomenal self-consciousness is pervasive in ordinary circumstances in which one is conscious – even on a restrictive interpretation of the relevant ‘ordinary circumstances’. If, by contrast, being cognitively self-conscious is *not* the only way in which we can be phenomenally self-conscious, then some version of (TSC) might be empirically accurate. To address this question, we must investigate whether there are determinates of phenomenal self-consciousness other than *cognitive self-consciousness*. In other words we must ask whether – as a matter of contingent empirical fact – there are ways of being phenomenally self-consciousness that are not instances of cognitive self-consciousness.

In this chapter, I will focus in particular on bodily experience, namely the experience of bodily sensations. In §6.1, I introduce the notion of non-conceptual *de se* content, and argue that if bodily experiences can have such content, then one can be phenomenally self-conscious – in a non-conceptual and non-cognitive way – in virtue of having a bodily

experience. In §6.2, I relate the question of whether bodily experiences can and do have non-conceptual *de se* content to a recent debate on the existence of a ‘sense of bodily ownership’. In §§6.3-6.4, I consider two empirical arguments in favour of the claim that some but not all bodily experiences involve such sense of bodily ownership. In §6.5, I argue that the second empirical argument supports the claim that ordinary bodily experiences involve a sense of bodily ownership, if this is understood as the claim that such experiences have non-conceptual *de se* content. Finally, in §6.6, I consider whether the hypothesis that bodily experiences normally involve non-conceptual *de se* content vindicates (TSC).

### 6.1 Non-conceptual *de se* content

As you are reading these lines, you are probably sitting in a chair. If so, bring your attention to what it is like for you to rest your back against the back of the chair. What this is like might include a sensation of *pressure* mediated by the activation of cutaneous mechanoreceptors, as well as a sensation of *warmth* mediated by the activation of cutaneous thermoreceptors. Upon feeling these bodily sensations, you should be in a position to make the following judgements:

- (a) There is pressure on my back.
- (b) There is warmth in my back.

If you make such judgements on the basis of attending to bodily sensations in your back, they are presumably *conscious* judgements. Furthermore, these judgements make use of the first-person concept, since they are suitably expressed by utterances containing the first-person pronoun ‘my’. Thus, self-ascriptions of bodily sensations like (a) and (b) are instances of conscious *de se* thinking. Since a subject engaged in conscious *de se* thinking is *ipso facto* phenomenally self-conscious, it follows that a subject consciously self-ascribing bodily experiences is *ipso facto* phenomenally self-conscious. However, this does not entail that a subject undergoing bodily experiences is *ipso facto* phenomenally self-conscious, without engaging in conscious *de se* thinking.

Bodily experiences do not have the same degree of cognitive sophistication as thoughts. It is very plausible that a number of non-linguistic creatures such as human infants and

non-human animals can undergo bodily experiences; but this presumably does not require such creatures to have the capacity to think, let alone to think of themselves as themselves. One important aspect of the difference between bodily experiences and thought is that the latter, but not the former, have concept-involving content. In particular, bodily experiences (e.g., the experience of warmth in one's back) require neither deployment – nor even possession – of the first-person concept that is a component of the content of *de se* thoughts. If being phenomenally self-conscious is a matter of being in a conscious mental state that makes use of the first-person concept – a conscious mental state with *conceptual de se* content –, then it follows that a subject undergoing a bodily experience is not *ipso facto* phenomenally self-conscious.

However, there are reasons to doubt that the intentional content of *all* conscious mental states need be concept-involving. Consider, once again, the mental states of non-linguistic creatures. Such creatures presumably represent the world as being a certain way. For example, a lion's visual experience of a gazelle grazing on a patch of grass might have an intentional content of the type <A gazelle is grazing on a patch of grass>, or perhaps simply <*This animal* is idle>. Having a mental state with such content is useful to guide the lion's behaviour, but it does not require the lion to possess and use the concepts GAZELLE, ANIMAL, GRAZING, OR BEING IDLE. Rather, it is plausible that the lion's perceptual state makes use of non-conceptual analogues of such concepts.<sup>1</sup> The same could be true of some conscious mental states that human have, such as perceptual or indeed bodily experiences.

If the intentional content of some mental states need not involve concepts, then one might wonder whether conceptual *de se* content is the only kind of intentional content that involves a representation of oneself as oneself, or whether there is a non-conceptual analogue of conceptual *de se* content. Thus, Bermudez (1998) and Peacocke (1992; 2014) argue that some mental states involve *non-conceptual de se* content. This hypothesis is attractive, because it can play an explanatory role in an account of *de se* thinking in two ways. First, as Bermudez (1998) suggests, it might be that a creature's capacity to engage in

<sup>1</sup>There is, of course, room for further disagreement about whether it is plausible that the intentional content of the visual experience of a lion would include a non-conceptual analogue of such a high-level concept as GAZELLE; but we can set this debate aside for the present discussion.

*de se* thinking (with *conceptual de se* content) is *developmentally* rooted in that creature's early-acquired capacity to undergo conscious experiences with *non-conceptual de se* content. Second, as (Peacocke 2014) suggests, the capacity to self-ascribe certain properties in conscious *de se* thought merely on the basis of undergoing certain conscious experiences might be explained by the hypothesis that the relevant conscious experiences have non-conceptual *de se* content, and that subjects can easily transition from the non-conceptual to the conceptual *de se*. This suggestion is particularly relevant for an account of bodily self-ascriptions: your capacity to judge that there is warmth in *your own* back, where this is a conscious *de se* thought formed merely on the basis of feeling warmth in your back, might be explained by the hypothesis that your bodily experience of warmth in your back has non-conceptual *de se* content. If this account is correct, then it seems to be the case that bodily experiences, such as the experience of warmth in one's back normally, have non-conceptual *de se* content. In Peacocke's words:

Some things are given in perception and in sensation in ways that do essentially involve the [non-conceptual] *de se*... A hand's being experienced as yours is part of the phenomenology of ordinary human experience. As elsewhere, this phenomenology should not be identified with any kind of judgement of a content 'that's mine'... It seems equally right to say that each normal human's experience of his or her body has the content that it is his or her own body.

Peacocke (2014, p. 51)

Nonetheless, as Peacocke (2014, pp. 30-9) emphasises, we can conceive of a subject whose conscious experiences entirely lack *de se* content (conceptual or not). The experiences of this subject would be, in Peacocke's terminology, at 'degree 0' of self-representation, rather than 'degree 1' – corresponding to the instantiation of non-conceptual *de se* content –, let alone 'degree 2' – corresponding to the instantiation of conceptual *de se* content. Furthermore, we can specifically conceive of a subject whose bodily experiences lack *de se* content. To borrow an example from Peacocke (2017, p. 292), it could be the case that a subject's proprioceptive experience of her leg as bent has the degree 0 content <*this leg* is bent>, rather than the degree 1 content <*my leg* is bent>, were 'my' expresses, not the first-person concept, but its non-conceptual analogue (that Peacocke calls 'first-person notion').

This raises the question of whether the bodily experiences we undergo in ordinary circumstances do have non-conceptual *de se* content at all. This question is directly relevant to the question we started with about phenomenal self-consciousness. Indeed, recall that I have defined phenomenal self-consciousness as follows:

(PSC) A subject *S* is *phenomenally self-conscious* at *t* if and only if (a) *S* is conscious of *S* [SELF] at *t*, and (b) *S*'s being conscious of *S* [SELF] at *t* makes a determinate constitutive contribution to *S*'s overall phenomenology at *t*.

If a subject *S* undergoes a bodily experience *e* with non-conceptual *de se* content (say, the content <*my back is warm*>), then *S* arguably meets this definitional requirement for phenomenal self-consciousness in virtue of having *e*. This is perhaps not obvious if we say that *S* is conscious of *her back as her own* in virtue of having *e*; but we might say, more appropriately, that *S* is conscious of *her back as a part of herself* in virtue of having *e*, or even that *S* is conscious of *herself as the bodily subject whose back is warm* in virtue of having *e*. Thus, to answer the question of whether a subject undergoing bodily experiences can be – and normally is – *ipso facto* phenomenally self-conscious, we need to determine whether bodily experiences can have – and normally do have – non-conceptual *de se* content; and, furthermore, we need an account of what it is for an experience to have non-conceptual *de se* content (at ‘degree 1’) rather than no *de se* content at all (at ‘degree 0’).

Consider once again what it is like for you to rest your back against a chair, and feel pressure and warmth on/in your back. These sensations of pressure and warmth are not, as it were, ‘free-floating’. One sense in which these sensations are not ‘free-floating’ has already been elucidated in chapter 1: they are undergone – enjoyed and endured – by some subject of experience. More specifically, they constitutively contribute to *your* overall phenomenology. But there is another sense in which these sensations are not ‘free-floating’: you do not feel pressure and warmth *simpliciter*, but you feel pressure and warmth *in your back*. In other words, these bodily sensations have a location, not merely in the sense that the cutaneous receptors causally mediating them happen to be located in a region of skin situated on your back, but in the sense that an aspect of *what it is like* for you to have such

bodily sensations is their being felt *in your back*, rather than in some other body part of yours, or in some region of space outside the boundaries of your own body.

This view of bodily experiences is not entirely uncontroversial. According to a more traditional conception of bodily experiences, these are just ‘raw feels’ that do not represent the world as being a certain way.<sup>2</sup> Perhaps there is some truth to this idea for a specific subset of bodily experiences. For example, some interoceptive experiences such as tiredness, nausea, thirst or hunger, and some vestibular experiences such as dizziness, might not always involve feeling something in a specific body part.<sup>3</sup> By and large, however, bodily experiences have at least spatial content. Suppose that, while resting your back against a chair, you also rest your right hand on a table in front of you; in doing so, you might have two distinct experiences of pressure: one caused by the stimulation of cutaneous mechanoreceptors in your hand, and another caused by the stimulation of cutaneous mechanoreceptors in your back. It is conceivable that the mechanoreceptors in both body parts might be stimulated in exactly in the same manner and to the same degree, such that your two experiences of pressure do not differ in intensity. Nonetheless, you should have no difficulty in distinguishing the two experiences on the basis of their phenomenology: one is a feeling of pressure on your right hand, and the other a feeling of pressure on your back. What it is like to feel pressure on one’s right hand differs from what it is like to feel pressure on one’s back.<sup>4</sup>

A plausible implication of the view that bodily experiences such as sensations of pressure or warmth are felt in a body part is that they constitute a form of consciousness of the relevant body part as feeling a certain way. Bodily experiences of this kind are not simply experiences *originating* from one’s body, but also experiences *of* one’s body. We can formulate this idea as the following *bodily consciousness principle* (BCP):<sup>5</sup>

<sup>2</sup>For example, “bodily sensations do not have an intentional object in the way perceptual experiences do” (McGinn 1996, p. 8).

<sup>3</sup>Armstrong (1962, p. 42) calls bodily experiences without spatial content *bodily feelings*, by contrast with *bodily sensations*. Of course, even if such experiences do lack *spatial* content, this does not entail that they are ‘raw feels’ without *any* intentional content, as the more traditional conception of bodily experiences would have it.

<sup>4</sup>See de Vignemont (2018, p. 68) for a similar example.

<sup>5</sup>I take it that (BCP) is commonly accepted in the philosophical literature as an implication of the view

(BCP) If a subject S feels a bodily sensation in a body part P, S is *ipso facto* conscious of (or *as of*) P.

(BCP) does not entail that a bodily experience consisting in feeling a sensation in a body part P has non-conceptual *de se* content. Indeed, paraphrasing Peacocke (2017)'s example, the content of an experience of warmth in one's back might simply be <*this back* is warm>, rather than <*my back* is warm>. In having an experience with such content, one would be *ipso facto* conscious of the back represented by the non-conceptual notion '*this back*'. But one would not be *ipso facto* conscious of that back *as one's own*, let alone conscious of *oneself as oneself*. The recent debate on the so-called 'sense of bodily ownership' is directly relevant to that distinction; in what follows, I will briefly present the debate, before considering several empirical arguments purporting to shed light on the *de se* component of bodily experiences.

## 6.2 The sense of bodily ownership

In recent years, a debate has emerged on whether bodily experiences ordinarily involve a 'sense of bodily ownership'.<sup>6</sup> Martin (1995) introduces this notion as follows:

When I feel an ache in my ankle, the ankle that feels hurt to me does not just feel like an ankle belonging to some body or other. Rather, the ankle feels to me to be part of my body... [I]n having bodily sensations, it appears to one as if whatever one is aware of through having such sensation is a part of one's body... This phenomenological quality, that the body part appears to be part of one's body [can be called] *a sense of ownership*...

Martin (1995, p. 269)

More recently, de Vignemont has extensively defended what she calls the *liberal view* of the sense of bodily ownership, according to which it is "a distinctive phenomenological quality in virtue of which one is aware of one's body as one's own" (de Vignemont 2018, p. 30). On the liberal view, when we have bodily experiences such as feeling warmth in

that some – perhaps most – bodily experiences involve feeling something in a body part. See for example Martin (1995): "[w]hen you feel an ache in your left ankle, it is your ankle that feels a certain way" (p. 268); "the experience [of feeling warmth in a body part] is as of a certain body part, as warm" (p. 269).

<sup>6</sup>For discussions of this claim, see for example Martin (1995), de Vignemont (forthcoming; 2007; 2013; 2018), Bermúdez (2005; 2011; 2017; 2018), Alsmith (2015), and Billon (2017).

one's back, feeling pain in one's knee, feeling pressure on one's hand, etc., "we [normally] have a primitive nonconceptual awareness of bodily ownership, which is over and above the experience of pressure, temperature, position, balance, movement, and so forth" (de Vignemont 2018, p. 13).<sup>7</sup>

De Vignemont contrasts the liberal view of the sense of bodily ownership with the *conservative view*, according to which there is no "distinctive experiential signature for the sense of bodily ownership: ownership is something that we believe in, and not something that we experience" (*ibid.*, p. 13). In the words of Bermúdez, cited as one of the main proponents of the conservative view:<sup>8</sup>

There are facts about the phenomenology of bodily awareness (about position sense, movement sense, and interoception) and there are judgments of ownership, but there is no additional feeling of ownership.

Bermúdez (2011, p. 167)

The nature of the disagreement between the liberal view and the conservative view of the sense of bodily ownership requires further elucidation. In particular, we need an account of what it means for the sense of bodily ownership to be an 'additional' to the phenomenology of bodily awareness (as Bermúdez puts it), or 'over and above' bodily sensations (as de Vignemont puts it). To a first approximation, the liberal view says the following: the phenomenology of the experience of pressure on your back is *not* exhausted by its being an experience of *pressure*, nor even by its being an experience of pressure *on a back* (or *on this back*). It is an experience of pressure *on your own back*. *Experiencing pressure on one's own back* is a determinate of the determinable *experiencing a body part as a part of one's own body*, which, for short, we can refer to as a determinable 'sense of bodily ownership'. Thus, on de Vignemont's liberal view, saying that a subject has a sense of bodily ownership at *t* does not mean that she has a *determinate* experience of ownership of her body (in general) at *t*. Rather, it means that she undergoes a *determinate* bodily experience in a specific body part (e.g., pressure on her back), and that what it is like to undergo that

<sup>7</sup>See also S. Gallagher (2005, pp. 28-29) and Kühle (2017). The existence of a sense of bodily ownership is often assumed in the growing body of empirical research on bodily awareness (see Tsakiris 2011 for a discussion).

<sup>8</sup>See also Alsmith (2015) and W. Wu (forthcoming).

determinate bodily experience is not exhausted by the bodily sensation (pressure) or its location (on the back), but also includes its being *on her own back*.

I take it that the gist of de Vignemont's view of the sense of bodily ownership can be translated in Peacocke's terminology as follows: bodily experiences that involve feeling a sensation in a body part normally have non-conceptual *de se* content, in so far as they represent (non-conceptually and non-cognitively) the relevant body part as a part of the subject's own body. Importantly, de Vignemont argues that this is not a necessary feature of bodily experiences, but only a feature of the ordinary bodily experiences of healthy subjects. Thus, on her view, one can in principle have a bodily experience that lack a sense of bodily ownership, that is, non-conceptual *de se* content. This allows de Vignemont to substantiate her liberal view with an examination of the phenomenal contrast between putative empirical examples of bodily experiences lacking a sense of bodily ownership, and ordinary experiences that have a sense of bodily ownership.

### 6.3 The argument from somatoparaphrenia

De Vignemont's strategy to defend her view of the sense of bodily ownership relies in large part on arguments from *phenomenal contrast*. Arguments from phenomenal contrast have a long history in philosophy, although the label was recently introduced by Siegel (2007). Such arguments purport to show that the best explanation of the phenomenal contrast between two experiences is the hypothesis that one involves a specific phenomenal feature F that the other lacks, as a way to arbitrate introspective disagreements about the existence of F. As we saw in the previous chapter, some arguments in favour of the liberal view of the phenomenology of thought are arguments from phenomenal contrast.<sup>9</sup>

De Vignemont's main argument from phenomenal contrast relies on clinical descriptions of the psychopathology known as somatoparaphrenia (de Vignemont [forthcoming](#);

<sup>9</sup>See Chudnoff (2015b), T. Horgan and Tienson (2002), Siewert (1998), and Strawson (1994).

2007; 2013; 2018).<sup>10</sup> Somatoparaphrenia is a monothematic delusion (typically caused by a brain lesion) characterized by the patients' belief that one of their body parts (usually contralateral to the side of their brain damage) is *not* really theirs. A number of patients also believe that the affected limb belongs to someone else. Here are a few characteristic reports from somatoparaphrenic patients cited in de Vignemont (2018, p. 21 and p. 216):

E: If this hand is not yours can I take it away with me? P: Of course! If you want it, I will give it to you as my gift, since I have no need for it. E: Do you want to move this hand away? Wouldn't you be sad without it? P: Yes, if it was mine, but it's not.

Invernizzi et al. (2013, p. 148)

How am I supposed to know whose hand is this? It's not mine.

Gandola et al. (2012, p. 1176)

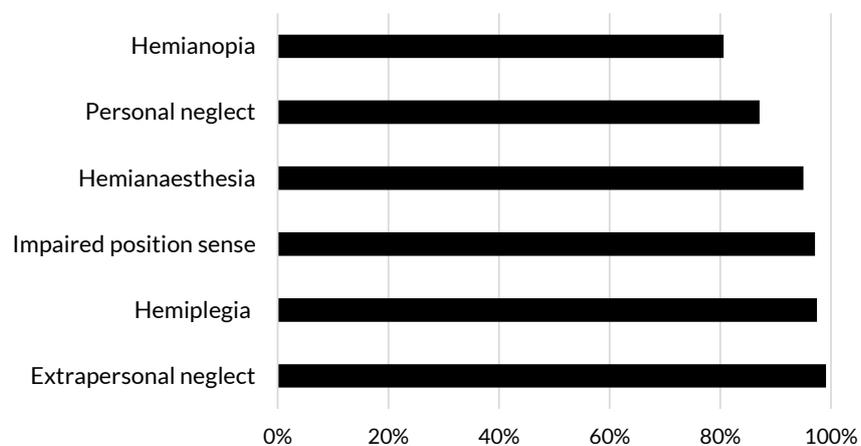
We can reconstruct de Vignemont's argument from phenomenal contrast as follows:<sup>11</sup>

- (P1) There is a phenomenal contrast between (a) what it is like for a somatoparaphrenic patient  $S_1$  to have bodily sensations in body part  $P_1$ , where  $P_1$  is the body part whose ownership  $S_1$  denies (e.g.,  $S_1$ 's right hand), and (b) what it is like for a healthy individual  $S_2$  to have bodily sensations in corresponding body part  $P_2$  (e.g.,  $S_2$ 's right hand).
- (P2) The best explanation of this phenomenal contrast, in light of reports from somatoparaphrenic patients, is that in having bodily sensations in  $P_1$ ,  $S_1$  is not conscious of  $P_1$  as a part of her own body, while in having bodily sensations in  $P_2$ ,  $S_2$  is conscious of  $P_2$  as a part of her own body.
- (C1) Therefore, for healthy individuals – but not for somatoparaphrenic patients – what it is like to have bodily sensations in a body part involves being conscious of that body part as a part of one's own body (i.e., having a sense of bodily ownership over that body part).

<sup>10</sup>In her rich book, de Vignemont (2018) relies on additional phenomenal contrasts between the ordinary bodily experience of healthy individuals and the bodily experiences associated with a number of other psychopathologies and bodily illusions to motivate her account of the sense of bodily ownership. For lack of space, I shall leave these aside in this chapter.

<sup>11</sup>The gist of this argument can be found in a more discursive format in de Vignemont (2013, pp. 648-650), de Vignemont (2018, pp. 18-23), de Vignemont (2019a) and de Vignemont (forthcoming).

In light of clinical evidence, the first premise of this argument calls for important qualifications. Somatoparaphrenia is associated with a large number of dramatic bodily and sensory disorders, including severe motor and somatosensory impairments (see Vallar and Ronchi 2009 and Romano and Maravita 2019 for reviews). Such impairments very frequently include unilateral neglect, both extrapersonal (inability to explore and represent the contralesional side of the extrapersonal space) and personal (inability to attend to sensory or motor aspects of the contralesional side of the body), hemiplegia on the contralesional side of the body (paralysis of half of the body, including the affected limb), impairment of the ability to determine the position of one's affected limb through proprioception, hemianaesthesia on the contralesional side (impairment or loss of tactile perception), and hemianopia on the contralesional side (loss of vision in half of the visual field). Figure 6.1 summarises the frequency of the bodily and sensory disorders most commonly associated with somatoparaphrenia, using data about a large sample of 131 patients pooled from 58 studies.



**Figure 6.1:** Bodily and sensory disorders commonly associated with somatoparaphrenia (adapted from Romano and Maravita 2019; percentages indicate the proportion of cases in which each impairment was observed in studies in which it was reported upon).

The first conclusion we can draw from these clinical data is that the bodily experiences of somatoparaphrenic patients, particularly on the contralesional side of the body, are very different from the bodily experiences of healthy individuals. So much so, in fact, that in many cases the ability of somatoparaphrenic patients to feel bodily sensations at

all in their affected limb is heavily impaired. For example, Halligan et al. (1995) report the following case:

Physical symptoms included left hemiparesis [muscle weakness and reduced motricity], left hemianopia, incontinence and impaired sensation. There was decreased muscle tone on the affected side and *gross impairment of sensation* on the left. *Light touch was absent throughout*; deep pressure was present but impaired. Joint position sense was *only present at the left hip* and there was astereognosis [inability to identify an object by active touch of the hands without other sensory input] in the left hand.

Halligan et al. (1995, p. 174, my emphasis)

This case report is far from isolated; for example, Romano and Maravita (2019) found that hemianaesthesia was observed in no less than 95% of cases in which it was reported upon – as indicated in figure 6.1. Thus, the differences between the bodily experiences of somatoparaphrenic patients and that of healthy individuals are so dramatic that it is not as easy as one might think to find a clear phenomenal contrast on which de Vignemont's argument can rely. Indeed, the argument should ideally focus on the contrast between two experiences involving the *same* type of bodily sensation in the *same* limb (e.g., touch on the right hand), to bring out a phenomenal feature that is missing in one experience and present in the other.

De Vignemont acknowledges that touch is frequently affected in somatoparaphrenia (2018, p. 40), but notes that many somatoparaphrenic patients can feel pain in the affected limb, and that in a handful of rare cases these patients can also feel and report touch in the affected limb (e.g., Bottini et al. 2002). Drawing on cases in which somatoparaphrenic patients have an intact ability to experience pain or touch in the affected limb, one might set up the argument by comparing, for example, what it is like for a somatoparaphrenic patient to feel pain or touch on/in her right hand, and what it is like for a healthy individual to feel pain or touch on/in her right hand. However, one might wonder what evidence we have that the determinate experience of pain (or touch) of the somatoparaphrenic patient and the determinate experience of pain (or touch) of the healthy individual differ, if at all, with respect to their phenomenal character. While there is ample evidence that *overall*, somatoparaphrenic patients have a range of abnormalities in the way they can experience their affected limb, there is less evidence to support the claim that patients with intact ability

to feel pain or touch in the affected limb have a different determinate phenomenology from healthy individuals when they experience pain or touch in the affected limb.

Consequently, the first premise of the argument does not seem to be, in effect, adequately supported by empirical evidence. Let us nonetheless suppose, for the sake of the argument, that there is indeed a clear phenomenal contrast not simply between the overall bodily experiences of somatoparaphrenic patients and healthy individuals, but in the specific experiences of, say, pain or touch that subjects in each category can undergo in a specific limb (corresponding to the affected limb of patients). The second step of the argument consists in arguing that in light of available reports, this phenomenal contrast is best explained by the hypothesis that only healthy individuals experience the relevant body part as a part of their own body. De Vignemont justifies this premise by examining the thematic content of the delusional reports of somatoparaphrenic patients – namely, the denial that the affected limb belongs to the patient’s body. As an example, she cites the following exchange between an clinician and a patient:

PATIENT: I still have the acute pain where the prosthesis is.

EXAMINER: Which prosthesis?

PATIENT: Don’t you see? This thing here. [*indicating his left arm*] The doctors have attached this tool to my body in order to help me to move... Once home could I ask my wife, from time to time, to remove this left arm and put it in the cupboard for a few hours in order to have some relief from pain?

Maravita (2008, p. 102)

Commenting on this case report, de Vignemont writes the following:

There is no doubt here that the patient was experiencing pain and that he was locating his pain in his left arm, and yet *there is also little doubt that he was not experiencing his left arm as his own*.

de Vignemont (2018, p. 40, my emphasis)

There is certainly little doubt that the patient of that case report was not *reporting* his left arm to be his own, and was *reporting* instead that what was actually his left arm was a prosthesis. From such reports, we can plausibly infer that the patient *believed* that what was actually his left arm was not his left arm but a prosthesis. But this is presumably not

sufficient to infer that the patient was not *experiencing* his left arm as his own, if this is intended to illuminate the *phenomenal* contrast between the patient's painful sensation in his left arm and a healthy individual's painful sensation in her left arm.

To motivate her interpretation of delusional reports from somatoparaphrenic patients, de Vignemont appeals to the two-factor theory of monothematic delusions:<sup>12</sup>

[A]ccording to the current most influential theory of delusion, the two-factor model, the thematic content of a particular delusion finds its origin in sensory or motor impairment leading to abnormal *experiences* that the patient tries to account for (Langdon and Coltheart 2000). This seems to be confirmed in somatoparaphrenia, in which the disownership delusion appears to be strongly anchored in abnormal feelings.

de Vignemont (2018, p. 24)

[According to the two-factor model], one needs to distinguish between the factors that trigger the initial implausible thought (and thus contribute to explaining the thematic content of a particular delusion), and the factors that explain the uncritical adoption of the implausible thought as a delusional belief (Coltheart, Langdon, et al. 2011; Langdon and Coltheart 2000). Abnormal rationality can only account for the feeling of confidence in the delusional beliefs, but not for their content.

de Vignemont (2018, p. 42)

According to the two-factor account, monothematic delusions are caused by two distinct factors rather than one: the first factor is supposed to explain the thematic content of the delusional idea or hypothesis, while the second factor is supposed to explain why the delusional idea or hypothesis is adopted and maintained as a belief (e.g., Coltheart, Langdon, et al. 2011; Davies et al. 2001). On this view, the first factor is the existence of abnormal data in need of an explanation, such as unusual perceptual data. Although a possible explanation for such data is provided by the delusional idea or hypothesis, the existence of the data is not sufficient in itself to explain why the delusional idea or hypothesis would be adopted as a belief – rather than being rejected – if cognitive processes were operating normally. Accordingly, the two-factor account postulates that a second explanatory factor is required, namely a deficit at the level of cognitive mechanisms responsible for the evaluation of explanatory hypotheses (possible explanations of the abnormal data). Thus, delusional patients come to adopt and maintain the relevant delusional belief because

<sup>12</sup>See also de Vignemont (2013, p. 649).

they fail to select the hypothesis that *best* explains the abnormal data, given what they otherwise know.

De Vignemont's appeal to the two-factor account calls for a few remarks. First, the hypothesis that the content of monothematic delusional beliefs is rooted in an anomalous experience is not specific to the two-factor account, but shared with the one-factor account offered by Maher (e.g., Maher 1999). According to the one-factor account, such anomalous experience is the single factor that explains the adoption of the delusional belief (*ibid.*, p. 551).

Second, the first factor of the two-factor account need not be a *conscious* experience. Admittedly, the first factor is often referred to as an *abnormal experience*; but the relevant notion of 'experience' is evidential rather than phenomenal. In other words, the first factor must be some abnormal data in need of explanation, regardless of whether the subject is conscious of these abnormal data. For example, Coltheart, Menzies, et al. (2010) emphasise that on the two-factor account of the Capgras delusion, characterised by the patients' belief that someone close to them has been replaced by an impostor, the first explanatory factor is lack of autonomic response to familiar faces. Since people are not conscious of the activities of their autonomic nervous systems, the lack of autonomic response to familiar faces is "not an abnormal experience, because it is not an experience" (*ibid.*, p. 264). The subsequent process of abductive inference that leads to the delusional belief need not enter into consciousness either:

Everything that preceded the occurrence of that belief and was responsible for the belief having come about – the stroke, the neuropsychological disconnection, the absence of an autonomic response when the wife is next seen, the invocation of a process of abductive inference to yield some hypothesis to explain this, and the successful generation of such a hypothesis – all of these processes are unconscious. What's conscious is only the *outcome* that this chain of processes generated: the conscious belief 'This person isn't my wife'.

Coltheart, Menzies, et al. (2010, p. 264)

Coltheart, Menzies, et al. (2010) also mention that this "this kind of analysis holds true for many forms of delusion, not just for Capgras delusion" (*ibid.*, p. 264). Of course, in the case of somatoparaphrenia, we have *independent* reasons to believe that patients may lack a number of normal *conscious* experiences that healthy individuals have, and

have a number of abnormal *conscious* experiences that healthy individuals do not normally have. Consequently, there is no doubt that there are phenomenal differences between the conscious experience of somatoparaphrenic patients, and that of healthy individuals; but we cannot rely on the two-factor account to make this claim.

Third, the claim that the thematic content of a delusion is causally mediated by the content of some anomalous experience does not entail that one should take patients' reports concerning their delusional *beliefs* as reports of the anomalous *experiences* in which the adoption of these beliefs are rooted (let alone *trustworthy* reports of such experiences). For example, somatoparaphrenic patients frequently report not only that the affected limb does not belong to them, but also that it belongs to someone else; and, in some cases, they even report that one of the doctor's limbs belongs to themselves (e.g., Hécaen and de Ajariaguerra 1952). It would be odd to claim that such reports are *best* explained by the hypothesis that these patients experience a distinctive feeling of the affected limb as belonging to the doctor, or a distinctive feeling of the doctor's limb as belonging to themselves.

With these remarks in mind, let us review briefly the two-factor account of somatoparaphrenia in hemiplegic patients provided by Coltheart, Langdon, et al. (2011, pp. 287-8). On this account, the first factor is the abnormal datum that a paralysed limb cannot be moved. Coltheart and colleagues suggest that patients can 'detect' this abnormal datum through preserved somatosensory and motor feedback when they attempt to move the affected limb. In turn, the cognitive deficits that constitute the second factor lead to the adoption of the delusional belief as the best explanation of the datum. In this particular case, it is natural to interpret Coltheart and colleagues as saying that patients with somatoparaphrenia *do* have a (conscious) bodily experience of motoric failure when they try to move the affected limb.<sup>13</sup> This two-factor account of somatoparaphrenia explains the thematic content of the delusion, because the idea or hypothesis that a limb is not one's own is an explanation of the datum that the limb cannot be moved by one's own will (not a *rational* explanation,

<sup>13</sup>Following Coltheart, Langdon, et al. (2011), this can be contrasted with patients suffering from anosognosia for hemiplegia, who are not able to register the abnormal datum that a limb cannot be moved because of the lack of somatosensory and motor feedback from that limb, leading them to adopt the delusional belief that they *can* move their arm.

but an explanation nonetheless). Importantly, this account does not and need not appeal to the putative lack of a sense of ownership in the patient's bodily experiences.

Romano and Maravita (2019, pp. 5-7) report on an interesting case that seems helpful to assess the relative plausibility of competing accounts of somatoparaphrenia. The case concerns a patient with severe motor impairment for the left arm, severe unilateral spatial neglect and personal neglect, and dense left hemianopia, but nearly intact tactile sensitivity, kinesthetic sensitivity (ability to detect when the affected limb is passively moved) and position sense (assessed by the ability to reproduce the posture of the affected limb with the unaffected limb on the other side of the body). What is interesting about this case is the fact that the patient could have bodily experiences such as tactile and proprioceptive experiences in her affected limb, as healthy individuals do. Consequently, it is in principle an ideal case to set up an argument from phenomenal contrast in favour of the existence of a sense of bodily ownership.

The following exchange between the patient and an examiner is reported:

*[After moving [the patient's] hand in front of her face, while she was keeping her eyes closed:]*

EXAMINER: Open your eyes, what is this?

PATIENT: This is my hand.

EXAMINER: Is this yours? Are you sure?

PATIENT: Yes, I feel that you are touching it.

EXAMINER: And if you could only look at it?

PATIENT: I don't know, it looks like it is mine.

*[After a few minutes [the examiner] moves [the patient's] left hand towards her face, after putting it on a pillow in order to reduce tactile cues:]*

EXAMINER: Now open your eyes. What is this?

PATIENT: A hand.

EXAMINER: Whose hand is this?

PATIENT: I don't know.

EXAMINER: Whose might it be?

PATIENT: It could be mine, but I don't really feel like it's mine. No it is not mine. It is swollen and it does not move, I don't feel like it is mine.

EXAMINER: So whose might it be?

PATIENT: It could be someone else's hand, but I don't know... I guess it's yours.

[Then, after spontaneously touching the left hand with the right one:]

PATIENT: No no, this is mine, now I feel it, I recognise it. It's mine!

Romano and Maravita (2019, p. 6)

As Romano and Maravita point out, the patient “could easily distinguish her hand when touched” (*ibid.*, p. 6): the delusional belief was only reported when tactile sensations were attenuated. This is *prima facie* difficult to explain within de Vignemont's account of somatoparaphrenia, according to which the delusional belief is rooted in the abnormal lack of a sense of bodily ownership over the affected limb (and perhaps a additional feeling of surprise that the affected limb does not feel like one's own). When the patient from Romano and Maravita's case report feels a tactile sensation in the affected limb, she immediately updates her delusional belief in favour of the correct belief that the limb is hers. Moreover, she is able to *justify* her belief that the limb is hers with reference to her tactile sensation in the limb (“I feel that you are touching it”, “this is mine, now I feel it”). By contrast, she justifies the delusional belief that the limb is not hers with reference to her inability to move it (“it does not move”). This is *prima facie* consistent with the account provided by Coltheart and colleagues.

Against the kind of account offered by Coltheart and colleagues', according to which somatoparaphrenia is rooted in the experience of motoric failure, de Vignemont points out that most individuals who are paralysed do not report lacking ownership of a limb (de Vignemont 2018, p. 169). However, this is what the two-factor account of somatoparaphrenia predicts. Indeed, the two-factor account holds that the first factor is not *sufficient* to cause the formation of the delusional belief. A second factor is needed to explain why this far-fetched belief is selected as the best explanation of the abnormal data. Thus, somatoparaphrenic patients should be expected to have an impairment of belief evaluation that non-delusional individuals with paralysis do not have, explaining why the latter do not come to believe that their paralysed limb does not belong to them.

In summary, the argument from phenomenal contrast that relies on available evidence regarding somatoparaphrenia does not conclusively suggest that such patients

lack a sense of bodily ownership that healthy individuals have. I will now introduce and discuss a new argument that focuses on the phenomenal contrast provided by certain drug-induced states.

#### 6.4 The argument from drug-induced disownership

Psychoactive molecules belonging to the pharmacological class of ‘classic psychedelics’ are known to produce dramatic alterations of conscious experience, which include in many cases alterations of bodily experience.<sup>14</sup> The subjective effects of classic psychedelics – such as mescaline, psilocybin, LSD and DMT – are mediated by agonism of serotonin 2A receptors.<sup>15</sup>

The molecule known as *N,N-Dimethyltryptamine* (DMT) has particularly interesting effects on bodily experience, as attested by a large number of online anecdotal reports. DMT is a psychedelic compound naturally occurring in a variety of plants such as *Mimosa tenuiflora*, with short-lasting (15 to 20 minutes) but powerful subjective effects. The short duration of DMT’s subjective effects makes it particularly convenient to use in controlled experimental studies. To address the limitations of anecdotal reports, I conducted in-depth interviews with the 15 participants of a randomised, placebo-controlled neuroimaging study on the effects of DMT (Timmermann et al. 2019). Participants were administered the drug intravenously while lying down with an eye mask, without being able to see their bodies. Soon after the effects of the drug had completely subsided, I conducted post-hoc interviews for one hour with each participant.<sup>16</sup>

All participants described an experience with a similar temporal structure. One of the first effects of the drug, a few seconds after administration, was the occurrence of salient bodily sensations of pressure, vibration and warmth all over their bodies. Participants described the gradual appearance of visual hallucinations of increasing complexity soon

<sup>14</sup>See Millière (2017) for a review.

<sup>15</sup>A given molecule acts as an *agonist* for a certain type of receptor if it fully activates the receptor that it binds to (somewhat like a key opening a lock), triggering cascade effects on brain activity and connectivity, which in turn may be associated with significant subjective effects.

<sup>16</sup>I used an interview technique known as microphenomenology, designed to obtain fine-grained descriptions of subjective experience while minimizing the risk of confabulation (Bitbol and Petitmengin 2017).

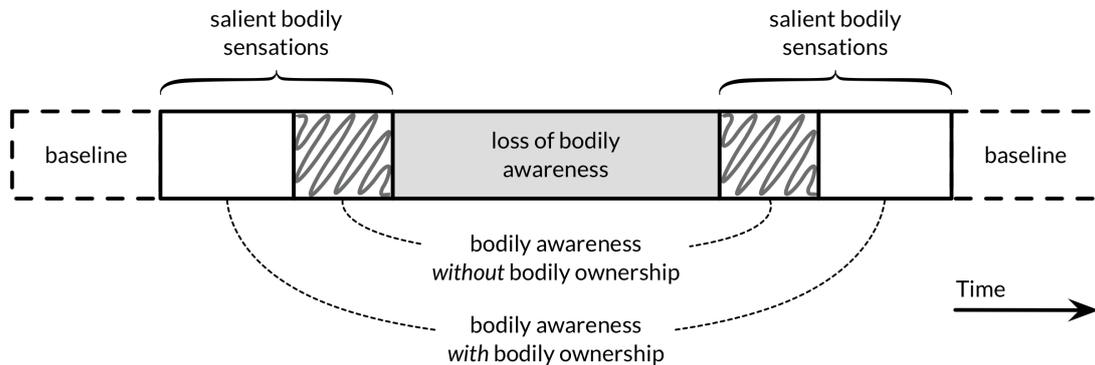


Figure 6.2: A model of the bodily effects of DMT

afterwards. As the visual hallucinations became immersive and three-dimensional a few minutes after administration, most participants reported losing any kind of bodily experience for several minutes, then gradually regaining bodily sensations when the effects of the drug started to wane.

Many participants described in great detail the transition phases leading respectively to the loss and retrieval of bodily sensations. Interestingly, these descriptions suggest that both transition phases included a brief period (ranging roughly from thirty seconds to a couple of minutes) during which participants had bodily sensations that they did not experience as in or on *themselves*. One possible interpretation of these reports is that participants initially experienced bodily sensations (e.g. pressure, vibration and warmth) in body parts that felt like part of their own body, then – during the first transition phase – these bodily sensations ceased to feel to them as if they were on/in their own body, and finally participants ceased to have any bodily sensation at the peak; and conversely when the effects of the drug started to subside, they progressively regained bodily sensations that – during the second transition phase – they did not experience on/in their own body, before they started experiencing again normal bodily sensations in body parts that felt like parts of their own body (figure 6.2).

Describing the initial transition phase leading to the loss of all bodily sensations, one participant (subject 6) mentioned having sensations in body parts that were not experienced as part of *herself*:

[T]his was kind of the intermediate point... and then in that moment [I remember] having a sensation in my toes or my hands but they didn't really feel like they were a part of me. And then I was just completely somewhere else and there was no sensation of body anymore, I completely lost that.<sup>17</sup>

Another participant (subject 12) described this transition as involving a bodily sensation – an intense sensation of pressure – but insisted that in experiencing this sensation she was not conscious of her own body as such:

I do feel the intense pressure on me. And I don't have any awareness of my body at all... [I]n every single cell of my body I feel pressure, but I didn't feel my body, I just felt pressure... I can't really say where I felt the pressure.

Several participants described in rich detail the parallel transition phase that occurred as they regained bodily sensations, insisting on the fact that these bodily sensations were not experienced in the usual way at first. Descartes famously wrote that “through these very feelings of pain, hunger, thirst, and so forth... I am not present in my body only as a pilot is present in a ship, but that I am very closely conjoined to it and, so to speak, fused with it, so as to form a single entity with it” (Descartes 1641/2008, p. 57). For a few minutes, one participant (subject 3) described bodily sensations as if he was, to use Descartes' metaphor, merely inside a vessel:

[T]hen what happened was I became aware of my body again... I had a body but I felt like a limp puppet... [I]t's like my body was just this kind of husk... So it's like a return to my body. But my sense of self was separate. That's interesting isn't it? So I guess I have this profound sense of my consciousness and sense of self being stuck in this kind of husk... The body is one thing, and me is something else... It felt like a sort of useless lump of flesh... It sort of felt like a vehicle that was carrying me around. I was aware that my self... was being carried around by this vehicle. And I could see it as a vehicle. It felt like a vehicle it felt like a machine in which I live sort of thing. It's like this is my house. This physical lump of flesh and my self... was something else... [I]t was sort of amusing that I was aware of my self... as residing in a body but dissociated from it.

The same participant elaborated on the metaphor of being inside a vehicle, to describe his experience of the bodily sensation of his arm being pinned as a doctor was taking blood samples:

<sup>17</sup>This report and the subsequent reports are taken from the unpublished transcripts of interviews I conducted with the participants of Timmermann et al. (2019).

Imagine if you've taken your car to the mechanics... and you had all these mechanics milling around, undoing the wheels, going under the bonnet, and you're in the car. It felt like that. [I was] sitting in my body as a sort of observer watching them mill around, pull bits... I felt as if I was in a machine, as opposed to [how one feels] normally... If you prod my arm today, now... you're prodding *me*. But then you weren't prodding *me*. You were prodding the machine. *I* was somewhere else. My soul was just in there.

Another participant (subject 11) also described the phase transition that occurred as he was regaining awareness of his body in terms that suggest that something was initially missing from the way in which he experienced bodily sensations:

Then the next stage would be when I was starting to become aware of myself again... it's kind of wrong to say 'I', because it wasn't like that, it was more like my body started to have sensations again...

These are a few excerpts sampled from 15 interviews, although a number of other segments describe similar experiences. Taken at face value, these reports describe a phenomenal contrast between (a) bodily sensations (e.g. of pressure or warmth) experienced in or on body parts that feel like parts of the subject's own body; and (b) bodily sensations that are experienced by the subject, yet not in or on body parts that feel like parts of the subject's own body (or on/in body parts that do not feel like parts of the subject's own body). On this interpretation, one might say that the bodily sensations in category (a) come with a 'sense of bodily ownership', while bodily sensations in category (b) lack a 'sense of bodily ownership'. Instead of using the terminology of ownership, one might say more specifically that bodily sensations in category (a) have non-conceptual *de se* content, while bodily sensations in category (b) do not. An alternative interpretation of the reports might take them to describe a different phenomenal contrast, between (a) bodily sensations (e.g. of pressure or warmth) experienced in or on body parts, and (b) bodily sensations experienced in or on body parts, and accompanied by an *additional* (and unusual) feeling of alienation or disownership with respect to the relevant body parts. On this alternative interpretation, reports do not support the hypothesis that bodily sensations on either side of the phenomenal contrast should involve a 'sense of bodily ownership' or non-conceptual *de se* content.

While available evidence might not allow us to arbitrate with high confidence in favour of either interpretation, there are nonetheless a few reasons why one might favour the first

one overall. Firstly, reports describe the transitions surrounding the peak of the experience in a symmetrical manner. The first transition is described as progressing through the *loss* of an ordinary feature of bodily experiences to the *loss* of all bodily experiences, while the second transition is described as progressing through the *retrieval* of bodily experiences, but lacking an ordinary feature, to the *retrieval* of this ordinary feature. This symmetrical structure is coherent with the first interpretation, according to which subjects first cease to experience bodily sensations as of their own body or body parts, before ceasing to experience bodily sensations altogether; and then first feel bodily sensations again that they do not experience as of their own body or body parts, before transitioning back to baseline and to normal bodily experiences as of their own body or body parts (figure 6.2).

Secondly, reports describe each transition as relatively linear. The first transition gradually progresses from salient bodily sensations to a complete loss of bodily sensations, while the second transition gradually progresses from the lack of bodily sensations to normal bodily sensations. If the second interpretation were correct, then neither of these transitions would be linear. The first transition would involve the emergence of an *additional* feeling of disownership accompanying bodily sensations before leading to the abrupt loss of both this additional feeling and the bodily sensations themselves. As for the second transition, it would progress from the lack of bodily sensations to the reappearance of bodily sensations plus an *additional* feeling of disownership, and then to the *loss* of this additional feeling. By contrast, the first interpretation fits well with the seemingly linear nature of the transitions described by the reports.<sup>18</sup>

## 6.5 From body to self

Let us come back to our initial question, namely whether a subject can be phenomenally self-conscious without engaging in conscious *de se* thinking – and, specifically, whether a subject can be phenomenally self-conscious by undergoing bodily experiences, without engaging in conscious bodily self-ascription. I argued that this would be the case if bodily

<sup>18</sup>Furthermore, this interpretation is also consistent with the relationship between the pharmacokinetics of the drug and the evolution of its subjective effects. For example, one might expect a gradual return to ordinary bodily experiences as the plasma concentration of the drug decreases.

experiences involved a non-conceptual form of self-representation – non-conceptual *de se* content – that partly determines what it is like for a subject to undergo them.

I suggested that reports from the DMT study provide some support for the existence of a phenomenal contrast between

- (a) bodily experiences with non-conceptual *de se* content, namely bodily sensations that are experienced on or in a body part P such that their subject is conscious of P as a part of their own body; and
- (b) bodily experiences without *de se* content, namely bodily sensations that are experienced on or in a body part P such that their subject is *not* conscious of P as a part of their own body.

Following Peacocke, one could schematically express the content of a sensation of warmth in one's hand belonging to the first category as <*my hand* is warm>, and the content of a sensation of warmth in one's hand belonging to the second category as <*this hand* is warm>.<sup>19</sup>

One question remains: how can we elucidate the difference between bodily experiences with non-conceptual *de se* content, and those that lack such content? In other words, what is required for a bodily experience to have non-conceptual *de se* content? For a bodily experience to have such content, it is not sufficient that it represents a sensation as in or on a body part, because otherwise there would be no difference between an experience with the content <*this hand* is warm> and an experience with the content <*my hand* is warm>. Peacocke (2017, p. 293) offers a sophisticated account of non-conceptual *de se* content that he calls the *agency-involving account*, according to which there are two severally sufficient and jointly necessary conditions for a non-conceptual component *c* of intentional content employed by a creature to be the non-conceptual *de se* component:

- (1) there is a range of action notions *A* for which the creature must be capable of being in mental states... with the content  
*c* is *A*-ing

<sup>19</sup>Some bodily sensations belonging to the second category might not be assigned to a body-part-specific location at all, in which case their non-conceptual content might simply be <*this body* is warm> or even <*this* is warm>.

where the state... is produced by the initiation of an *A*-ing by the reference of *c*;  
and

- (2) there is a range of notions *F* of bodily properties, spatial properties, and past tense properties *F* such that the creature is capable of being in mental states... with the content *c* is *F*; where in these attributions,

*c* is *F*

is accepted (in central basic cases) if and only if

this body is *F*

is also accepted.

Peacocke (2017, p. 293)

This account is offered as an elaboration on the account given in Peacocke (2014), according to which the non-conceptual analogue of the first-person concept (the ‘first-person notion’) is individuated by the reference rule that on any occasion of its occurrence in a mental state, it refers to the subject of that state. Conditions (1) and (2) of the agency-involving account are meant to clarify how this reference rule secures the reference to the subject, rather than something else.

On the agency-involving account, a subject will be capable of having a bodily experience with non-conceptual *de se* content of the type <*my hand* is warm> if and only if: by condition (1), *S* is capable of being in mental states produced by *S*’s initiating some action with her hand that represent the subject’s hand as performing that action; and by condition (2), the content <*my hand* is warm> is accepted if and only the content <*this body* is warm> is also accepted. On the face of it, the agency-involving account seems to predict, somewhat implausibly, that paralysed subjects who cannot move a given body part should not be capable to have bodily experiences with *de se* content in that body part. Yet it seems that subjects can have bodily experiences in a paralysed body part (sensation of pressure or warmth, for example), where these experiences do not feel different from similar experiences in non-paralysed subjects.<sup>20</sup>

Peacocke (2019) briefly mentions that the agency-involving account would allow for a paralysed subject to keep using the first-person concept in thought, because she “would

<sup>20</sup>For example, Gooch et al. (1993) report the case of 12 patients with prolonged paralysis but no sensory loss after treatment with neuromuscular junction blocking agents. See also de Vignemont (2018, p. 169).

still know what it is like to act” (p. 127). However, he does not explicitly mention whether such a paralysed subject would be able to have bodily experiences with *non-conceptual de se* content. It is certainly implausible that the subject’s capacity to undergo such experiences in some body part should depend upon their *knowledge* of what it is like to act with that body part. If this is indeed what the agency-involving account predicts, then it seems overly intellectualistic. Furthermore, subjects in the DMT study did presumably have the knowledge of what it is like to act, and their bodily experiences satisfied the first condition of the agency-involving account; yet they seemed to lack *de se* content altogether.

For these reasons, we may tentatively propose an alternative account of the non-conceptual *de se* content of bodily experiences.<sup>21</sup> On this account, a bodily experience on or in a body part P has non-conceptual *de se* content *only if* it has not only a body-part-specific location, but also an egocentric location specified with respect to the subject’s trunk and determined at least in part by proprioceptive information.<sup>22</sup> Thus, a tactile sensation in one’s right hand is normally not merely represented as occurring in *this hand*, but is also represented as occupying a relative location *to the right of the subject’s body*.<sup>23</sup> In the DMT study, the bodily experiences that subjects reported during the transition phases appeared to lack egocentric location, although some of them appeared to have body-part-specific locations. This is not particularly surprising, given that classic psychedelic drugs like DMT have significant disrupting effects on multisensory integration, including integration of bodily signals.<sup>24</sup> In the relevant phase of the DMT-induced state, it appears that bodily sensations are no longer properly integrated with proprioceptive information;

<sup>21</sup>In her (2018), de Vignemont also proposes an alternative account of the sense of ownership, the *affective account*, according to which an experience of a bodily sensation involves a sense of ownership *only if* the body part in which the sensation is located is represented within the ‘protective body map’, a spatial representation of the body that grounds protective behaviours. I cannot do justice to this account here, although I note that there are reasons to prefer an account that does not link the *de se* content of bodily experiences to the disposition to engage in protective behaviours (see Bradley 2019 for a discussion).

<sup>22</sup>I am less confident that this is a sufficient condition for a bodily experience to have non-conceptual *de se* content.

<sup>23</sup>On the distinction between these two kinds of bodily location, see de Vignemont (2018, pp. 71-3) and Bermúdez (2017).

<sup>24</sup>For a review and discussion, see Letheby and Gerrans (2017) and Millière (2017).

as a result, they no longer feel connected to a self, and the body parts in which they occur are no longer represented as parts of the bodily subject.

Reports from the DMT study also suggest that the terminology of ‘ownership’ is not really adequate to capture the non-conceptual *de se* content that appears to be present in ordinary bodily experience and missing in the bodily experiences one might have after intoxication. As de Vignemont (2018) mentions, the term ‘sense of bodily ownership’ is unfortunate, because “I do not ‘own’ my body [as I] own my laptop, my flat, and my books” (pp. 1-2). Interestingly, participants of the DMT study did not extensively use the terminology of ownership (or disownership); rather, they reported having bodily sensations in body parts that did not feel like a part of *themselves* (“didn’t really feel like they were a part of *me*”), or more generally that the body in which they experienced bodily sensations did not feel like *themselves* (“*The body* is one thing, and *me* is something else”). Similarly, it might be slightly misleading to describe ordinary bodily sensations experienced on/in a body part as involving consciousness of that body part *as one’s own* or *as a part of one’s own body*. A more appropriate way to describe the non-conceptual *de se* content of such experiences might be to say that in feeling a bodily sensation in body part P, one is normally – non-conceptually and non-cognitively – conscious of P *as a part of oneself*, or better yet that one is – non-conceptually and non-cognitively – conscious *of oneself* as the bodily subject whose body part P one feels the sensation on/in.<sup>25</sup>

The upshot of this discussion is that in ordinary circumstances – in the sober, wakeful state – a subject feeling a bodily sensation in a body part is *ipso facto* phenomenally self-conscious in a non-conceptual and non-cognitive way. This is not a constitutive feature of all bodily experiences, because there are unusual circumstances (e.g., in DMT intoxication) in which bodily experiences are not instances of phenomenal self-consciousness. Nonetheless, it answers the first question we set out to address: being *cognitively* self-conscious –

<sup>25</sup>Note that this account need not commit us to the “strong metaphysical claim” that “I am my body” (de Vignemont 2019b). Rather, the claim is that in experiencing a bodily sensation in one’s body part, one is normally conscious of oneself as a bodily (or ‘embodied’) subject. This is why the relevant experiences provide us with a way of “gaining knowledge of ourselves *as physical and spatial things*” (Evans 1982, p. 220; see also Cassam 1997, chapter 1).

engaging in conscious *de se* thinking – is not the *only* way in which one can be phenomenally self-conscious. There is a distinct determinate of phenomenal self-consciousness that we might call *bodily self-consciousness*.

## 6.6 Bodily self-consciousness in ordinary experience

I have already suggested that bodily self-consciousness is not ubiquitous in all conscious experience, let alone constitutive of consciousness. By definition, one can only be bodily self-conscious when one experiences bodily sensations; and, furthermore, there appear to be cases in which one can have bodily sensations without being bodily self-conscious (e.g., in DMT intoxication). Recall, however, that the overarching purpose of Part II is to assess (TSC), the claim that we are phenomenally self-conscious *in ordinary circumstances*. I have argued so far that one can be phenomenally self-conscious by engaging in conscious *de se* thinking, or by undergoing bodily experiences with non-conceptual *de se* content. In the previous chapter, I suggested that conscious *de se* thinking is far from ubiquitous in ordinary experience; we can now ask what is the prevalence of bodily experiences with non-conceptual *de se* content in ordinary experience. In the set of ‘ordinary circumstances’ to which (TSC) is restricted, do we always have, as William James (1890) puts it, a “feeling of the same old body always there” (p. 242)?

Let us, once again, consider a very restrictive understanding of ‘ordinary circumstances’, limiting (TSC) to the ordinary wakeful (and sober) experience of healthy, neurotypical adult human beings. How often, in such circumstances, do such subjects experience bodily sensations in any body part? It is difficult to give a specific answer to this question. It is fairly plausible that one does not *constantly* experience bodily sensations in any specific body part in particular – for example in one’s hand or foot. Interestingly, Schwitzgebel (2007) used experience sampling to assess how frequently participants wearing a device beeping at random intervals had tactile experiences just before the beeping sound. One group of participants had to report on the occurrence of tactile experience in general (in any body part), while the other had to report on the occurrence of tactile experience in their left foot specifically. Answers were divided into three categories, ‘yes or leaning yes’, ‘undecided/don’t know’, and ‘no or leaning

no' (p. 19). Schwitzgebel found a high variability in the answers given by participants in the left foot group, with one participant reporting tactile experience in their left foot ('yes or leaning yes') 92% of the time, and another 16% of the time. The median for the left foot group was 49%. The median proportion of 'yes or leaning yes' answers for participants who had to report on their tactile experience in general was significantly higher at 76.5%, with less variability among participants.

Taken at face value, these data do not support the claim that – in the ordinary wakeful state – we *constantly* have tactile experiences, left alone tactile experiences in a specific body part such as one's left foot. Admittedly, bodily experiences are not limited to tactile experiences; the kind of bodily sensations we experience in body parts also include sensations of muscle stretch, tendon tension, joint position, pain and temperature. But Schwitzgebel (2007) notes that he defined 'tactile experience' broadly for participants to include "any tactile, somatic, nociceptive, or proprioceptive experience" (p. 22).

De Vignemont herself seems to take a more conservative stance on the prevalence of bodily experience and the sense of bodily ownership in ordinary circumstances:

Like most bodily awareness, the phenomenology of ownership is recessive, staying at the margin of consciousness... [W]e have very little experience of [the] bodily self. Although we receive a constant flow of information about our body, we are most of the time barely aware of our body, let alone of our body as our own.

de Vignemont (2018, p. 22)

However, while de Vignemont suggests that bodily experiences are not something we normally *attend to*, she does not explicitly deny that bodily experiences are constantly present in the ordinary wakeful state, albeit 'at the margin' of attention (see in particular de Vignemont 2018, p. 11). Ultimately, it is difficult and perhaps impossible to assess with great specificity how prevalent bodily experience is. Extrapolating from Schwitzgebel's data, we can tentatively suggest that in the ordinary wakeful state, healthy and neurotypical adult human beings undergo bodily experiences on/in body parts *most of the time* – although such experiences might not be salient. On the assumption that bodily experience on/in body parts normally have non-conceptual *de se* content, it would follow that healthy and neurotypical adult human subjects are bodily self-conscious, and hence phenomenally self-conscious, in *most* of the ordinary wakeful state. Nonetheless, there might also be times in

the ordinary wakeful state during which healthy and neurotypical adult human subjects are *neither* cognitively self-conscious, *nor* bodily self-conscious – perhaps, for example, when they are entirely focused on a demanding task or absorbed in movie, without paying any attention to their bodies.

## 6.7 Conclusion

In this chapter, I have argued that, aside from the conceptual *de se* content of *de se* thoughts, some mental states might have non-conceptual *de se* content. I subsequently argued that bodily experiences usually have such content, and that their having such content contributes to what it is like to undergo them. A subject undergoing a bodily experience with non-conceptual *de se* content is *ipso facto* phenomenally self-conscious, without having to engage in conscious *de se* thinking. While very frequent in ordinary experience, however, it is difficult to say how pervasive bodily experiences with non-conceptual *de se* content really are. Accepting the existence of bodily self-consciousness alongside cognitive self-consciousness certainly gets us closer to vindicating (TSC), but it is not quite sufficient. In the next chapter, I will argue that there is yet another distinct determinate of phenomenal self-consciousness whose pervasiveness in ordinary experience provides better support for (TSC).

# 7

## Spatial Self-Consciousness

Ask yourself what it is that you see hiding the surroundings as you look out upon the world – not darkness surely, not air, not nothing, but the ego!

Gibson (1979, p. 112)

In the previous chapters, I have argued that a subject undergoing a bodily experience with non-conceptual *de se* content is *ipso facto* phenomenally self-conscious. Consequently, phenomenal self-consciousness is not reducible to cognitive self-consciousness. In this chapter, I will argue that there is yet another way in which a subject can be phenomenally self-conscious, namely by undergoing perceptual experiences with non-conceptual *de se* content.

Being able to navigate through one's environment is fundamental for the survival of many organisms endowed with the capacity for locomotion. Spatial navigation crucially involves locating oneself with respect to one's surroundings, and thus representing spatial properties of environmental landmarks conveyed by sensory input. Some perceptual experiences – particularly those associated with distal sensory modalities such as vision and audition – seem to represent the locations of perceived objects with respect to the location of the organism. For example, a particular object might be represented as being located at a certain distance to the right of the perceiving subject.

Perceptual experiences that have this kind of spatial content represent the world *from somewhere*, namely from a specific vantage point constrained by the location and anatomical configuration of the subject's perceptual apparatus. We might say that such experiences are *perspectivally structured*, where this means to a first approximation that their spatial content is organised in a way that is sensitive to the location, orientation and anatomy of

the subject's body. If the relevant experiences represent the locations of environmental landmarks relative to the location of the subject's body, then there is a sense in which they must conversely represent the location of the subject's body relative to the locations of environmental landmarks. A number of authors have dwelled upon this observation to suggest that perspectively structured perceptual experiences involve a basic form of self-consciousness that consists in experiencing one's location (with respect to one's perceived environment) *as one's own*:

The *ecological self* is the self as perceived with respect to the physical environment: 'I' am the person here in this place... Are we *conscious* of our ecological selves?... I believe... that it is often accompanied by a definite – and often powerful – kind of awareness.

Neisser (1988, pp. 36-41)

[I]n egocentric spatial perception the objects of perception are experienced as standing in spatial relations *to the perceiver*... Egocentric spatial perception can therefore be described as self-locating; in experiencing objects as spatially related to one, one literally experiences the bodily self as located in the perceived world.

Cassam (1997, pp. 52-53)

Consider... the everyday case in which an ordinary person forms a belief with the content 'I am in front of a door', and does so for the reason that he sees a door ahead of him. His visual experience represents the door as bearing a certain spatial relation to him. This is so even if he cannot see or otherwise experience his own body on this particular occasion.

Peacocke (1998, p. 264)

Perceptual content... can vary along a *perspectival* dimension, in regard to how things *look* (or appear) from the vantage point of the perceiver... This corresponds to the fact that perception is, at once, a way of keeping track of how things are, and also of our relation to the world. Perception is thus world-directed and self-directed.

Noë (2005, p. 168)

[S]imply in virtue of its perspectival character, visual experience can include the location of the perceiver among its face value contents... [V]isual experience can be self-locating even when 'the self' is entirely out of view: that is, even when the perceiver's body is nowhere in the field of vision.

Schwenkler (2014, p. 139)

In what follows, I will assess the claim that perspectively structured perceptual experiences can (and generally are) instances of phenomenal self-consciousness. In §7.1, I define the notion of perspectival spatial content more precisely. In §7.2, I argue that an experience

can in principle have perspectival spatial content without *de se* content. In §§7.3-7.4, I discuss two arguments from phenomenal contrast purporting to show that some but not all perspectivally structured perceptual experiences have non-conceptual *de se* content. First, I examine an argument drawing upon the contrast between the experience of watching a scene in person and the experience of watching the same scene on a video, and argue that it is unconvincing (§7.3). Second, I introduce a novel argument drawing upon the contrast between experiences induced by immersive and non-immersive virtual reality respectively, and argue that it provides compelling support for the claim that perspectivally structured perceptual experiences normally have non-conceptual *de se* content (§7.4). In §7.5, I draw upon this empirical argument to provide an elucidatory account of what it takes for a perceptual experience to have non-conceptual *de se* content. Finally, in §7.6, I ask whether this account vindicates the typicalist claim that conscious subjects are phenomenally self-conscious in ordinary circumstances (TSC).

## 7.1 Perspectival spatial content

Focus your attention on your current visual experience. This experience is not chaotic and undifferentiated, but organised or structured in several respects. One respect in which your visual experience is structured is that it represents objects as unified wholes, rather than meaningless arrangements of colours and shapes. But your visual experience is also spatially structured with respect to a specific perspective or vantage point. Indeed, it represents the location of objects or landmarks in your environment within a subject-centred perspective, also called an egocentric frame of reference: objects are seen as being at a certain distance (near or far) and in a certain direction from you (left, right, up, down or directly in front). In other words, your visual experience represents the world *from somewhere*. This description applies not only to visual experience, but also to auditory experience.<sup>1</sup>

A *frame of reference* refers to a system of coordinates used to represent locations in space, together with a set of reference points that uniquely fix coordinates within that frame.

<sup>1</sup>See Evans (1982, p. 155). This is consistent with empirical evidence about the processing of auditory stimuli: “early auditory cortex primarily represents sound source location relative to ourselves” (Town et al. 2017, p. 1).

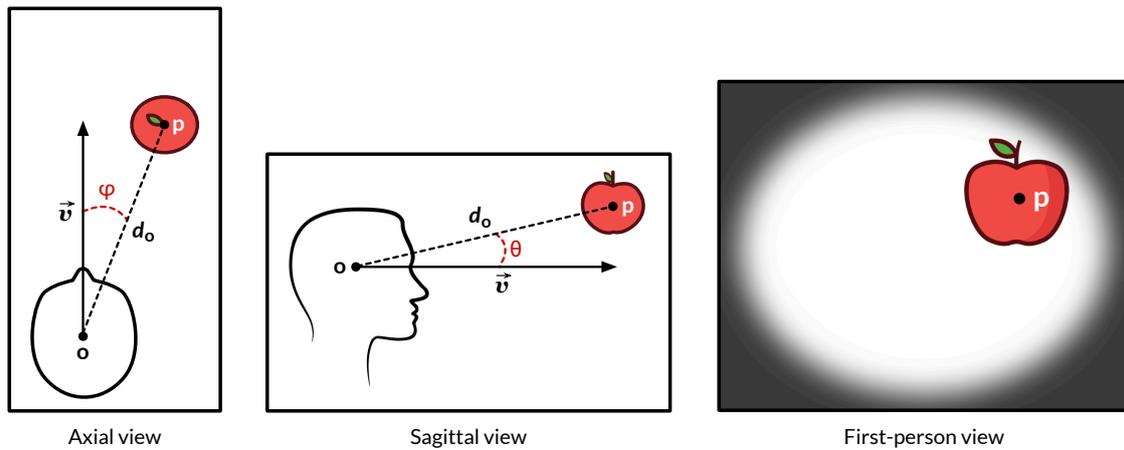


Figure 7.1: Representation of a location in an egocentric frame of reference.

An *egocentric* frame of reference specifically refers to a system of coordinates in which locations are represented with respect to a particular perspective specified by a subject-specific point of origin (also called ‘ego’) and a subject-specific reference axis (also called ‘orientation’) (Klatzky 1998). In the context of perception, this point of origin is the particular vantage point of a perceiving subject, roughly determined by the physical configuration of the relevant sensory organs (e.g., the eyes), and the reference axis is the direction in which the perceiving subject is facing.

Locations in an egocentric frame of reference can be formally characterised using a spherical coordinate system as follows (see figure 7.1):<sup>2</sup> a point  $p$  has coordinates  $(d_o, \varphi, \theta)$  defined with respect to the origin  $o$ , where  $d_o$  is the egocentric distance of the point (distance between  $p$  and  $o$ ),  $\varphi$  is the egocentric bearing of the point on the axial plane (angular deviation of the vector from  $o$  to  $p$ , referenced on a left-right axis from the orientation  $\vec{v}$  of the subject), and  $\theta$  is the egocentric bearing of the point on the sagittal plane (angular deviation of the vector from  $o$  to  $p$ , referenced on an up-down axis from the orientation  $\vec{v}$  of the subject).

Egocentric frames of reference are contrasted with allocentric frames of reference, in

<sup>2</sup>I use a spherical coordinate system here because this is arguably the most intuitive way to define locations as they are represented by perceptual experiences, but it is possible to use a cartesian coordinate system instead (the formalisms are interchangeable, although there might be computational benefits in representing locations within one system of coordinates rather than the other).

which locations are represented within a system of coordinates that is independent of the subject's particular vantage point and location. Intuitively, allocentric frames of reference can be said to be 'map-like', in so far as they represent locations independently of the subject's position and orientation, just as maps represent the locations of landmarks independently of the user's position and orientation.

There is converging empirical evidence that humans and many non-human animals represent locations both in egocentric and allocentric frames of reference. For example, it is now well-established that rodents and primates (among other mammals) make use of a 'cognitive map' representing both the spatial configuration of their environment in an allocentric frame of reference and their own location in that environment (O'Keefe and Nadel 1978; Tolman 1948). This cognitive map is encoded by specialised neurons, in particular 'place cells' in the hippocampus that fire whenever the animal is at specific locations in the environment (O'Keefe and Dostrovsky 1971), and 'grid cells' in the medial entorhinal cortex that fire in a regular hexagonal lattice of locations tiling the entirety of the environment independently of the animal's location (Hafting et al. 2005). Intuitively, grid cells can be considered to provide the allocentric coordinate system of the cognitive map, while place cells pin down the location of the animal in this coordinate system (similarly to the 'you are here' sign on a printed map).

Having a cognitive map storing allocentric representations of locations is important for animals to find their way from one position to another, by representing their trajectory independently from their particular point of view at any one time. However, the allocentric representation of locations on a cognitive map must be derived from a representation of locations in an egocentric frame of reference, since the spatial content of perception initially involves only the second type of representation. Indeed, given that sensory organs are located on the animal's body, the locations of environmental landmarks are represented within an egocentric (body-centred) frame of reference before they can be represented in an allocentric (world-centred) frame of reference. Consequently, it has been suggested that there are neural mechanisms transforming the short-term egocentric representation of locations given by the spatial content of perception into long-term allocentric representations for spatial navigation, which are subsequently converted back into an egocentric

frame of reference prior to behavioural output (Bicanski and Burgess 2018; Byrne et al. 2007; Hinman et al. 2019).

As we have seen, representing the location of a landmark in an egocentric frame of reference requires information about (a) its *egocentric distance*, or distance between the relevant landmark and the point of origin of the perceiving subject's perspective, and (b) its *egocentric bearing*, or angular deviation between the vector from the origin to the landmark and the orientation of the perceiving subject. With respect to visual perception, the egocentric bearing of landmarks can be computed from the location of their inverted image on the retina. For example, in the illustration of figure 7.1, the location of the retinal image of the apple should provide enough information in principle to represent the apple as located *to the upper left* of the perceiving subject's orientation, where this orientation is given by the vector from the eye to any location in the centre of the visual field.<sup>3</sup>

In turn, egocentric distance of visual landmarks is computed through a variety of depth cues, some of which are given directly by static features of visual stimuli (pictorial depth cues), while others depend on additional information (nonpictorial depth cues). Pictorial depth cues include occlusion (nearer objects can cover up parts of farther objects, providing information about their relative distance), relative size (nearer objects form a bigger retinal image, while farther objects form a smaller retinal image, providing absolute distance information if the size of objects is known), and height in the visual field (the egocentric distance of an object can be computed as a function of the perceiver's eye height and the angle between the line of sight to the horizon and the line of sight of the subject). In turn, nonpictorial depth cues include motion parallax (when the subject moves, the retinal images of stationary objects move on the side of the direction of movement in the inverse proportion to their egocentric distance), muscular feedback from accommodation

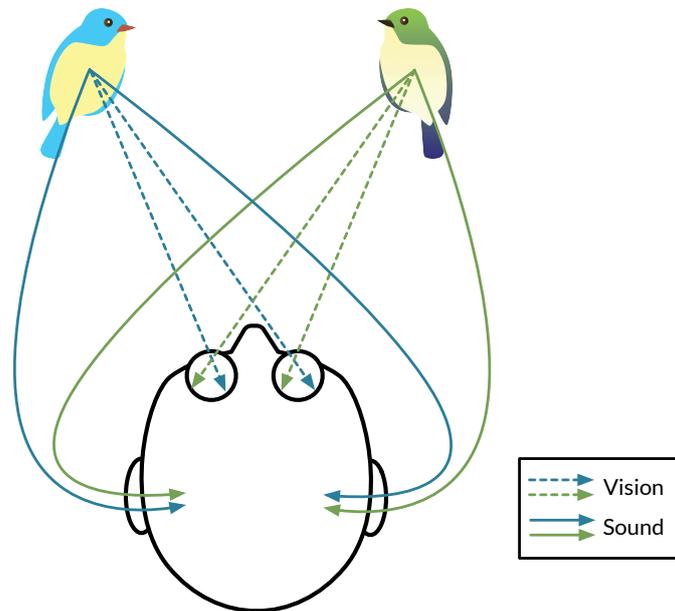
<sup>3</sup>This is a simplified account, since the orientation of the eyes can be misaligned with the orientation of the head and/or the orientation of the torso. The egocentric bearing of visual landmarks is not only computed with respect to a retinotopic (eye-centred) frame of reference, but also with respect to a head-centred frame of reference and a trunk-centred frame of reference. Information about the retinal location of visual stimuli is not sufficient to compute the egocentric bearing of a visual landmark in a head-centred or trunk-centred frame of reference, which requires additional information about the orientation of the eyes with respect to the orientation of the head, and the orientation of the head with respect to the orientation of the torso (respectively). Such information can be provided by proprioceptive input about joint position and vestibular input about head movement.

(the change in the curvature of the lens of the eye allowing it to focus on objects at various distances), muscular feedback from convergence (the angle between the optical axes of the eyes increases when the subject looks at nearer objects), and most importantly binocular disparity (the difference between the retinal locations of visual stimuli in the left eye and in the right eye, see figure 7.2).<sup>4</sup>

The auditory system also uses a variety of depth cues to compute the egocentric distance of the source of sounds, including the relative loudness of familiar sounds and the amount of reverberation. Egocentric bearings are less straightforward to compute for auditory stimuli than for visual stimuli, since there is no equivalent of the spatial layout of retinal images for sounds. Nonetheless, they can be computed mainly through binaural disparities, such as differences in the timing and loudness of auditory stimuli between the left ear and the right ear. Indeed, if a sound wave arrives at the head from one side, it has to travel further to reach the far ear than the near ear. This difference in pathlength results in a time difference between the sound's arrivals at the ears, as well as a difference in intensity between the sound registered by each ear, which are detected to compute the direction of the sound's source (figure 7.2).

Beyond the static representation of locations in an egocentric frame of reference, audiovisual perception can also provide dynamic information about the movement of the organism. Thus, during locomotion, patterns of optic flow in the visual field specify the speed and heading direction of the subject (a phenomenon dubbed 'visual kinesthesia' by Gibson 1979). Importantly, the heading direction of a moving subject need not coincide with the reference axis of the egocentric frame of reference of perceptual content. For example, a crab's heading direction is normally at a 90° angle from the reference axis (orientation) of its visual reference frame. Likewise, human beings can move in a direction other than the one in which they eyes are oriented. Consequently, the subject's heading direction can be computed in an egocentric frame of reference on the basis of audiovisual

<sup>4</sup>There is a further distinction between monocular and binocular depth cues. Depth cues are monocular if they can be computed from the information provided by a single eye, while they are binocular if they require information from both eyes. All pictorial depth cues are monocular. However, nonpictorial depth cues can be either monocular (such as those provided by motion parallax and accommodation) or binocular (such as those provided by convergence and binocular disparity).



**Figure 7.2:** Detection of egocentric locations by the visual and auditory systems through left-right disparities.

input, in addition to the subject's orientation.<sup>5</sup> In so far as the content of audiovisual perception does represent heading direction during locomotion, this is yet another feature of egocentric spatial content.

The contents of visual and auditory perception thus have broadly perspectival features, representing locations within an egocentric frame of reference. The perspectival structure of perceptual content specifies the subject's location, orientation and heading direction relative to her perceived environment. In what follows, I will assess whether the subject of a conscious perceptual state with perspectival spatial content is *ipso facto* self-conscious.

## 7.2 *de hinc* and *de se* content

The empirical evidence reviewed in the previous section suggests that visual and auditory experiences typically have perspectival spatial content structured by an egocentric frame

<sup>5</sup>This is not to say that heading direction can be computed from audiovisual input *independently* from the subject's orientation, since the latter constitutes the reference axis of the egocentric frame of reference within which heading direction is initially represented (prior to the transformation of egocentric coordinates into allocentric coordinates).

of reference. Furthermore, it suggests that such experiences represent spatial relations between environmental landmarks and the point of origin of this frame of reference, whose location roughly coincides with the location of the subject's head in ordinary circumstances. However, there is a *prima facie* difference between

- (1) a perceptual experience representing locations as being at a certain distance and in a certain direction from *here* – where *here* refers to the location of the point of origin of the frame of reference structuring the perceptual field (roughly coinciding with a point in the middle the subject's head, behind the eyes); and
- (2) a perceptual experience representing locations as being at a certain distance and in a certain direction from *me*.

While (b) involves *de se* content, namely a representation *as of oneself*, (a) involves only *de hinc* content, namely a representation of the location of the experience's point of origin in relation to surrounding landmarks that does not represent *oneself* as occupying that location.<sup>6</sup> The empirical evidence reviewed so far does not entail that perspectively structured experiences should involve *de se* content, as opposed to merely *de hinc* content.

The distinction between *de hinc* content and *de se* content relates to Peacocke (2014)'s distinction between three degrees of self-representation. A perspectively structured perceptual experience with merely *de hinc* content would involve what Peacocke calls 'degree 0' of self-representation: it would represent the locations of landmarks with respect to *here*, where *here* denotes the point of origin of its frame of reference. It is plausible that the intentional content of perceptual experiences, like that of bodily experiences, need not involve concepts. Consequently, a perceptual experience with *de hinc* content would employ the non-conceptual analogue of the *de hinc* concept *HERE*. By contrast, a perceptual experience involving what Peacocke calls 'degree 1' of self-representation would have non-conceptual *de se* content rather than merely non-conceptual *de hinc* content: it would represent the locations of landmarks with respect to *me*, where *me* denotes the subject

<sup>6</sup>I borrow the label '*de hinc* content' from Schellenberg (2016), as it is a convenient way to characterize this kind of content by analogy with *de se* content. While '*de se*' means 'about oneself' in Latin, '*de hinc*' means 'about here'.

as such. Peacocke considers the possibility that a simple organism could have perceptual experience involving only ‘degree 0’ representations:<sup>7</sup>

At what we can call Degree 0, the subject does not enjoy mental states with *de se* contents. The subject itself really is an element of reality, but it does not represent itself as such... There could be a creature, let us take it to be a spherical underwater creature, whose perceptual apparatus is positioned in such a way that it cannot perceive its own body. It is moved passively through the fluid. It enjoys perceptions of objects and events around its location... This creature remains at Degree 0, however, because it never represents anything as standing in certain relations to itself. None of its perceptual states have *de se* contents of such forms as *that thing is that direction from me*. Rather, they have *here*-contents [or *de hinc* contents], such as *that thing is that direction from here*.

Peacocke (2014, p. 30)

By claiming that a creature limited to ‘degree 0’ of self-representation is conceivable, Peacocke suggests that perspectively structured experiences without *de se* content are at least metaphysically possible, if not nomologically possible.<sup>8</sup> Schellenberg also argues that perspectively structured experiences enjoyed by creatures as sophisticated as humans could have merely *de hinc* content without *de se* content:<sup>9</sup>

[T]he thesis that one perceives objects in relation to one’s location does not imply that one perceives objects to *one’s* right or to *one’s* left. It implies only that one perceives objects to the right or to the left. So the idea that one perceives objects in relation to one’s location does not depend on being aware of oneself as standing in spatial relations to those objects.

Schellenberg (2007, p. 620)

Note that neither Peacocke nor Schellenberg denies that some perspectively structured experiences do have *de se* content; rather, they both suggest that having perspectival spatial content is not *sufficient* for having *de se* content, and consequently that something more

<sup>7</sup>See also Campbell (1994, p. 119): “it seems absolutely clear visual proprioception [i.e., visual experience providing information about the subject’s location, orientation and movement] is possible for creatures that are not self-conscious and have no grasp on the first person.”

<sup>8</sup>In his 2014, Peacocke falls short of endorsing the claim that experiences at ‘degree 0’ of self-representation actually exist. In more recent publications, however, he acknowledges that “there is a much wider range of important and interesting examples at Level 0 than I [previously] considered”, and that “there is a wide range of conscious states and events that a subject can enjoy, including perceptual experience... without employing the *de se*” (Peacocke 2016, p. 352-3; see also Peacocke 2017, p. 290).

<sup>9</sup>See also Schellenberg (2016, p. 341).

must be said to vindicate the claim that some perspectively structured experiences represent the location of the subject as such.

For a conscious experience to have non-conceptual *de se* content is, as Peacocke (2014, p. 51) puts it, “part of the phenomenology” of that experience. Any conscious experience with *de se* content involves conscious self-representation, and consequently a subject undergoing a conscious experience with *de se* content should be *ipso facto* phenomenally self-conscious. This is true for bodily experiences, and it is also true for perceptual experiences. Thus, if some perspectively structured perceptual experiences have non-conceptual *de se* rather than merely *de hinc* content, this should be reflected in what it is like to have such experiences.

In what follows, I will defend the claim that some but not all perspectively structured experiences have non-conceptual *de se* content by using the same strategy as that of the previous chapter. I will first discuss an argument from phenomenal contrast in favour of this claim drawn from previous discussions of perspectival spatial content, and argue that it is ultimately unconvincing. I will subsequently put forward a new argument from phenomenal contrast appealing to recent empirical evidence from psychological research on virtual reality. I will argue that the relevant phenomenal contrast cannot be explained by the occurrence of conceptual *de se* thinking, and is best explained by the hypothesis that in normal circumstances, perspectively structured experiences have non-conceptual *de se* content.

As in the case of bodily experiences, this empirical argument can be supplemented with a philosophical elucidatory account of what it takes for a subject to have perspectively structured experiences with non-conceptual *de se* content. I previously argued that Peacocke’s *agency-involving account* was not completely compelling as an explanation of a subject’s capacity to undergo bodily experiences with non-conceptual *de se* content, because the constitutive role of action-capabilities in this account seems too restrictive in the case of bodily experiences. In what follows, I will argue by contrast that an account in the vicinity of Peacocke’s *agency-involving account* can explain a subject’s capacity to undergo perspectively structured perceptual experiences with non-conceptual *de se* content rather than merely *de hinc* content.

### 7.3 The argument from film-watching

Drawing on Campbell's remarks about "the position of someone watching a film" (Campbell 1994, p. 120), Alsmith (2017) suggests that watching a film is an example of a conscious experience with perspectival spatial content but no *de se* content. Here is how Alsmith discusses the example of a film-watcher called Johanna:

Film is a medium for the presentation of perspectival images of scenes. Film images can and typically do have both *limitation structure* and *egocentric structure*. Johanna's film experience, her experience of a filmed scene through film images, will also be structured in these ways. Unless one is involved in the creation of the film, the filmed scene will typically not be anywhere in one's whereabouts. Hence, Johanna can experience a filmed scene that is not itself in her locale and be aware of this fact. And given that she is aware of this fact, *ex hypothesi* we then have no reason to think that Johanna would represent the perspective from which she experiences the filmed scene as where she herself is located. Film experience thus seems to be a candidate case in which one can experience something from elsewhere, without representing oneself as being in that location.

Alsmith (2017, p. 269)

Alsmith does not explicitly frame his use of this example as an argument from phenomenal contrast. Rather, he appeals to this example to motivate the claim that some perceptual experiences with perspectival spatial content do not represent the location of their egocentric frame of reference's point of origin *as the subject's own location*. However, we can modify his example slightly to make it more suitable for an argument from phenomenal contrast. Suppose that Johanna takes a video of an exotic bird with her phone while on vacation – e.g., she films the bird from the position in which she is standing, roughly at eye level, while looking at the bird rather than her phone's screen. Call her visual experience of the bird 'bird-watching'. Now suppose that a few weeks later, Johanna watches this vacation film on her television at home. Call her visual experience of the film 'film-watching'.<sup>10</sup> Alsmith's remarks should apply to Johanna's film-watching experience: even though Johanna was

<sup>10</sup>While it is not strictly necessary that the film attempts to capture Johanna's bird-watching experience as faithfully as possible for the two experiences to be contrasted, their similarity does narrow down in principle the difference in phenomenal character between them, which facilitates the argument from phenomenal contrast. Nonetheless, as I argue below, this pair of experiences is far from being a 'minimal pair', namely a pair of experience whose phenomenal contrast is *entirely* due to the fact that one of them lacks *de se* content.

involved in the creation of the film, the filmed scene is not anywhere in her whereabouts while she watches the film at home, and she is aware of that fact.

An argument from phenomenal contrast could then proceed as follows:

- (P1) There is a difference in phenomenal character between Johanna's bird-watching experience and her film-watching experience, which are both perspectively structured perceptual experiences.
- (P2) This difference in phenomenal character is best explained, at least in part, by the hypothesis that Johanna's bird-watching experience involves non-conceptual *de se* content while her film-watching experience does not.
- (C1) Therefore, some but not all perspectively structured perceptual experiences involve non-conceptual *de se* content.

While the first premise of this argument is beyond dispute, the second premise is problematic for several reasons. Firstly, the experience of watching a film differs in many ways from the experience of actually watching the scene captured by the camera as it unfolded. Some of these differences are due to properties of the film itself. Alsmith is correct in pointing out that film images do contain enough information to represent locations in an egocentric frame of reference, whose point of origin is the location of the camera itself. In particular, films contain pictorial depth cues (e.g. occlusion and relative size) that are often sufficient to estimate egocentric distance from the camera's location, and their spatial layout straightforwardly conveys information about the egocentric bearings of objects in the scene (e.g., an object is located *to the right of the camera*).

Unlike the output of the visual system, however, the output of a camera is monoscopic (i.e. obtained from a single viewpoint) and not sensitive to the viewer's eye movements; consequently, films lacks nonpictorial depth cues from convergence, accommodation and binocular disparity. In other words, film images are 'flat', and they are impossible to mistake for unmediated visual experiences of the world if only for that reason. Consequently, it is difficult to assess whether the presence or lack of non-conceptual *de se* content should be part of the best explanation of the phenomenal contrast between Johanna's bird-watching

and film-watching experiences. This is the case even if Johanna took good care of filming the scene from an anatomically congruent perspective roughly coinciding with the actual location of her head.<sup>11</sup>

Secondly, there is a significant difference between the content of the visual experience of watching a film, and the content of the relevant film.<sup>12</sup> The visual experience of watching a film on a screen does not merely represent the content of the film, but also the subject's environment (e.g., the screen itself, the lights, or any other environmental landmark in the room). In particular, the experience of watching a film represents locations of the subject's actual environment in an egocentric frame of reference. For example, Johanna's film-watching experience should represent the location of the television screen as being *in front* of her location, just as Johanna's bird-watching experience represents the location of the bird as being *in front* of her location. Thus, the point of origin of her film-watching experience is not the location of the camera relatively to landmarks depicted in the film, but Johanna's actual location in the room while she is watching the film.

Consequently, there is no reason to think that Johanna's film-watching experience differs from her bird-watching experience with respect to the way in which it represents spatial relations between the environmental landmarks and her own location. At best, one might say that her film-watching experience *embeds* a secondary egocentric frame of reference corresponding to the scene depicted on the film, whose point of origin does correspond to the location of the camera with respect to the filmed scene. However, this is not sufficient to conclude that Johanna's film-watching experience is an example of a perceptual experience without *de se* content.

Thirdly, according to Alsmith, it follows *ex hypothesi* from the fact that Johanna is aware that the location of the filmed scene is not her actual location (while she is watching the film) that she does not "represent the perspective from which she experiences the filmed scene as where she herself is located" (Alsmith 2017, p. 269). A natural way to

<sup>11</sup>One could even imagine a futuristic scenario in which Johanna has a retinal implant that directly records the input of one eye and accurately reconstructs Johanna's visual experience as a video. The resulting film would lack depth cues from binocular disparity all the same.

<sup>12</sup>There is further question about whether the notion of *content* implicated in each case is the same, or similar enough to warrant the analogy. I will leave that question aside here.

interpret the antecedent claim is that Johanna *knows* that her actual location (say, in her apartment in Oxford) differs from the location in which the bird was filmed (say, in the Amazonian rainforest). However, it is implausible that such abstract knowledge would impact the content of Johanna's visual experience, unless one endorses a strong version of the claim that perceptual experience is cognitively penetrable.<sup>13</sup> Consequently, Alsmith's claim threatens to prove too much.

For these reasons, the argument from film-watching does not provide sufficient motivation for the claim that some but not all perspectively structured perceptual experiences have *de se* content. I will now argue that a more compelling argument from phenomenal contrast can be developed by substituting virtual reality for film as the medium of the contrasted scenarios.

#### 7.4 The argument from virtual reality

The most compelling arguments from phenomenal contrast rely on pairs of experiences whose phenomenal character is as similar as possible, while nonetheless differing in some respect. As we have seen in the previous section, the argument from film-watching is unconvincing because the contrasted experiences differ significantly with respect to their phenomenal character. Consequently, there are many ways in which one could explain what the phenomenal contrast between these experiences consists in, and it is unclear that the best explanation should appeal to the phenomenology of self-location.

Virtual reality (VR) offers a way to simulate the experience of perceiving a scene much more faithfully than watching a video on a flat screen. VR broadly refers to any technology allowing users to perceive a virtual (computer-generated) environment in three-dimensions. A typical VR system consists of a stereoscopic head-mounted display connected to a computer, and broadcasting images and sounds from the virtual environment to the user's eyes and ears in such a way that they have an illusory perceptual experience of the virtual environment in three-dimensions (figure 7.3).

<sup>13</sup>I will argue in the following section that one's visual experience can represent the location of its point of origin as one's own location even though one *knows* that it is not one's actual location.

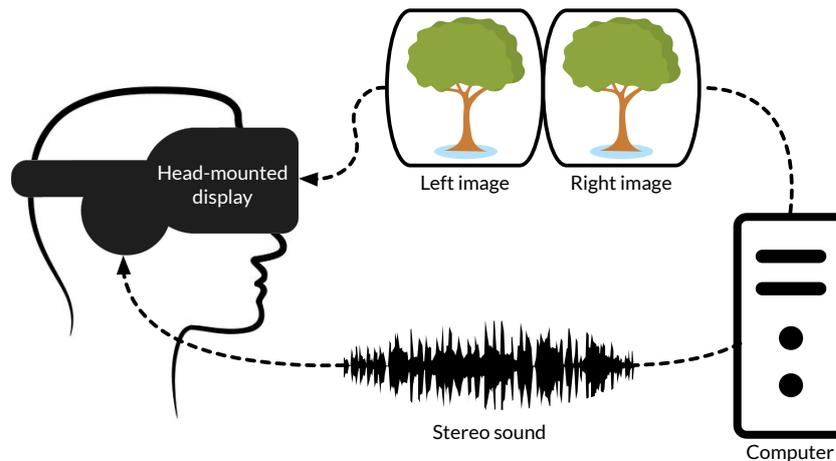


Figure 7.3: A typical VR system using a computer and a stereoscopic head-mounted display.

Visual experiences provided by VR differ from the visual experience of watching a film in several respects. Firstly, because the user has a head-mounted display strapped directly to her face, the images of the virtual environment broadcast to each eye through a lens cover almost the entirety of the user's field of view, as opposed to a small region of the user's field of view for a film watched on a flat screen at a comfortable distance. Secondly, head-mounted displays are stereoscopic, meaning that they broadcast a slightly different perspective on the virtual environment to each of the user's eyes, to generate the illusion of perceiving this environment in three dimensions. Consequently, the general spatial features of the visual experience that one has in VR are very similar to those of a veridical visual experience. In particular, VR-induced experiences have at least *de hinc* content, in so far as they represent the locations of virtual environmental landmarks in an egocentric frame of reference, whose point of origin is given by the location from which the scene is digitally rendered within the virtual environment.

There is a further technical distinction between immersive and non-immersive VR systems.<sup>14</sup> The former differ from the latter in that they not only enable the user to have a perceptual experience of a virtual environment, but also deliver "the ability to perceive

<sup>14</sup>I follow Slater (2009) in understanding 'immersive' as a technical property of VR systems that depends upon specific features such as head-tracking, rather than a phenomenal property of experiences induced by VR systems.

through natural sensorimotor contingencies” (Slater and Sanchez-Vives 2016, p. 5). Indeed, immersive VR systems allow users to interact with the virtual environment, by tracking the user’s bodily movements in three dimensions and adjusting the graphical input to the head-mounted display accordingly. Specifically, such systems include rotational and positional head-tracking, meaning that they detect rotations of the user’s head on longitudinal, lateral and vertical axes (roll, tilt and yaw), as well as changes of the head’s location in world-centred coordinates (forward/backward and leftward/rightward). Thus, the visual perspective broadcast into the head-mounted display is rendered in real time by the computer to match the user’s head movements. For example, if the user rotates their head to the right, the visual perspective rendered in the head-mounted display will be rotated to the right at the same time, as if the user was actually present in the virtual environment and looking around the scene.

There is a large body of empirical evidence suggesting that immersive VR can induce a ‘sense of presence’ in the virtual environment, unlike non-immersive VR. This experience, also known as ‘place illusion’, refers to the feeling of ‘being there’ within the computer-generated environment that one perceives through the head-mounted display (Heeter 1992; Held and Durlach 1992; Slater 2009). The empirical literature on the sense of presence in immersive VR suggests that it is related to “perceived self-location” (Wirth et al. 2007, p. 497), namely to the representation of the egocentric location specified by the user’s perspective within the virtual environment *as the user’s own location*.

The relevant empirical evidence supports an argument from phenomenal contrast according to which the best explanation of the fact that perceptual experiences induced by immersive VR involve a ‘sense of presence’, unlike perceptual experiences induced by non-immersive VR, is that the former have non-conceptual *de se* content (Peacocke’s ‘degree 1’ of self-representation), while the latter merely have non-conceptual *de hinc* content (Peacocke’s ‘degree 0’ of self-representation):

- (P1) There is a difference in phenomenal character between the perspectively structured perceptual experience of a given virtual environment induced by immersive VR and non-immersive VR respectively.

- (P2) This difference in phenomenal character is best explained, at least in part, by the hypothesis that the perceptual experience induced by immersive VR has non-conceptual *de se* content while the perceptual experience induced by non-immersive VR has merely non-conceptual *de hinc* content.
- (C1) Therefore, some but not all perspectively structured perceptual experiences have non-conceptual *de se* content.

In the empirical literature on the sense of presence in VR, subjective reports are generally collected through dedicated questionnaires. Van Baren and IJsselsteijn (2004) found that no less than 28 different questionnaires have been used to measure the sense of presence in virtual environments. The most widely used questionnaires are the Witmer and Singer Presence Questionnaire (PQ; Witmer and Singer 1998), the Slater-Usoh-Steed Questionnaire (SUS; Usoh et al. 2000), and the Igroup Presence Questionnaire (IPQ; Schubert et al. 2001) – the IPQ being the most reliable (Schwind et al. 2019).

Several studies contrasted scores on the IPQ in non-immersive VR – involving a stereoscopic head-mounted display *without head-tracking* – and in immersive VR – involving a stereoscopic head-mounted display *with head-tracking*. Thus, Regenbrecht and Schubert (2002) assessed the sense of presence with the IPQ in two conditions: in the passive condition, participants were shown a pre-recorded sequence in the virtual environment from a first-person perspective, while in the self-movement conditions, the participant's movements were tracked in real time so that they could freely move their point of view within the virtual environment. The authors found that self-movement significantly increased the sense of spatial presence as measured by the IPQ.

A more recent study tested the influence of head-tracking on the sense of presence by comparing the experience of riding a virtual rollercoaster in VR without and without tracking the participants' head movements (T. L. Y. Wu et al. 2019). Scores on the IPQ were higher in the head-tracking condition; specifically, the IPQ items "Somehow I felt that the virtual world surrounded me" and "I felt present in the virtual space" were rated significantly higher in the head-tracking condition compared to the passive condition.

The phenomenal contrast between perceptual experiences induced by immersive VR and non-immersive VR respectively cannot be explained by the hypothesis that the former, but not the latter, have perspectival spatial content. Indeed, both immersive and non-immersive VR systems use a stereoscopic head-mounted display that provides more than enough visual and auditory depth cues – including strong cues from binocular and binaural disparities – for the resulting experience to represent the locations of virtual landmarks within an egocentric frame of reference. A more plausible hypothesis to explain the phenomenal contrast is that perceptual experiences induced by immersive VR have non-conceptual *de se* content, while perceptual experiences induced by non-immersive VR only have *de hinc* content. This is in line with reports of the ‘sense of presence’ as the experience of being *oneself* located within the virtual environment.

There are, however, three alternative interpretations of the relevant phenomenal contrast:

- (a) **Cognitive interpretation.** The ‘sense of presence’ reported in immersive VR does not pertain to the phenomenology of perceptual experiences, but to the subject’s judgements or beliefs about their location.
  - (b) ***Something missing* interpretation.** Perceptual experiences in immersive VR do not involve a phenomenology of self-location that perceptual experiences in non-immersive VR lack; rather, the immersive VR experiences lack an abnormal feeling that is an aspect of the non-immersive VR experiences.
  - (c) ***Something else* interpretation.** Perceptual experiences in immersive VR do have some phenomenal property (or properties) that perceptual experiences in non-immersive VR lack; but the additional phenomenal property (or properties) of immersive VR experiences have nothing to do with their having non-conceptual *de se* content.
- (a) The cognitive interpretation is easy to dismiss. In so far as participants in VR experiments are not delusional, it would be ludicrous to suggest that *post hoc* self-reports regarding the ‘sense of presence’ in immersive VR reflect genuine judgements or beliefs

that they were actually located within the virtual environment while they were wearing the headset. Mel Slater, one of the pioneers of psychological research on the effects of immersive VR, eloquently articulates the implausibility of this interpretation:

Of course no one, not even when they are standing by a virtual precipice with their heart racing and feeling great anxiety, ever *believes* in the reality of what they are perceiving. The whole point of presence is that it is the *illusion* of being there, notwithstanding that you know for sure that you are not. It is a perceptual but not a cognitive illusion, where the perceptual system, for example, identifies a threat (the precipice) and the brain-body system automatically and rapidly reacts (this is the safe thing to do), while the cognitive system relatively slowly catches up and concludes 'But I know that this isn't real'. But by then it is too late, the reactions have already occurred.

Slater (2018, p. 432)

(b) Refuting the *something missing* interpretation is less straightforward. Indeed, it is not implausible that some experiences with immersive VR might lack a phenomenal feature instantiated by some experiences with non-immersive VR. In particular, stereoscopic head-mounted displays *without head-tracking* cause a mismatch between visual input and proprioceptive and vestibular cues about self-motion, since the viewpoint rendered by the computer is not sensitive to the user's head movements.<sup>15</sup> Such mismatch can be associated with an unpleasant combination of symptoms known as 'cybersickness' – a special case of motion sickness (M. Gallagher and Ferrè 2018). Furthermore, there is some evidence that cybersickness is negatively correlated with scores on various presence questionnaires (Weech et al. 2019).

These findings are consistent with the *something missing* interpretation: non-immersive VR would involve a feeling of cybersickness that is missing in immersive VR experience, because the latter matches sensory input to self-motion through head-tracking. Accordingly, one might suggest that reports of the 'sense of presence' in immersive VR are indicative of the *lack* of cybersickness when head movements are tracked, rather than an additional phenomenology of self-location.

However, this hypothesis does not stand up to careful scrutiny of available evidence. First, it should be noted that there is significant inter-individual variability in susceptibility

<sup>15</sup>The visual system contains neurons that are specifically attuned to this kind of mismatch (Keller et al. 2012; Zmarz and Keller 2016).

to cybersickness; around 20% of people do not experience cybersickness even with low quality VR systems (Cobb et al. 1999). If the self-reported ‘sense of presence’ was merely indicative of a lack of cybersickness, we should expect 20% of participants to score high on presence questionnaire with both immersive and non-immersive setups, which is not the case.

Second, the existence of a negative correlation between cybersickness and the sense of presence is still a matter of disagreement, with a number of studies suggesting instead either a positive or a null correlation (see Weech et al. 2019 for a review and discussion). In any case, the existence of a negative correlation between cybersickness and the sense of presence should not be surprising if the former is mediated by multisensory conflict between visual cues and proprioceptive/vestibular cues while the latter is mediated by multisensory congruency between these cues.

Thirdly and most importantly, subjective reports themselves suggest that the ‘sense of presence’ in immersive VR is not merely related to the absence of cybersickness, but also to a positive change in phenomenology. Items from questionnaires designed to measure the sense of presence in virtual environment are not related to the lack of unpleasant symptoms, but to the feeling of being spatially present within the environment (e.g., “In the computer generated world I had a sense of ‘being there’”, “I felt present in the virtual space”, and “Somehow I felt that the virtual world surrounded me” in the IPQ). There is also a large amount of anecdotal evidence regarding the sense of presence from online discussions between users of consumer-oriented VR headsets with head tracking (mainly used for immersive video games). Many VR users describe their first experience with immersive VR as involving the feeling of being present in the virtual environment:

My first experience [with immersive VR] was running the tutorial. You’re in a huge room, and it feels like you’re really there... It doesn’t feel like the screen is changing in front of your eyes, it actually genuinely feels like you’re just looking around the room.<sup>16</sup>

I was blown away at how fast my brain just accepted it... I really felt like I was ‘there’.<sup>17</sup>

<sup>16</sup>[www.reddit.com/r/Vive/comments/7g550v/i\\_know\\_its\\_been\\_asked\\_a\\_million\\_times\\_but\\_whats](http://www.reddit.com/r/Vive/comments/7g550v/i_know_its_been_asked_a_million_times_but_whats).

<sup>17</sup>[www.reddit.com/r/virtualreality/comments/9k2u8o/first\\_time\\_in\\_vr](http://www.reddit.com/r/virtualreality/comments/9k2u8o/first_time_in_vr).

[W]hen I put the [VR headset] on last night and tried this game I almost cried the moment I flew towards and then entered an enormous space station and looked around. This was real, I was there. Disbelief was completely and utterly suspended.<sup>18</sup>

These reports do not seem to describe the mere absence of cybersickness; in fact, many anecdotal online reports from VR users highlight that the sense of being present within the virtual environment can be concomitant with a sensation of sickness. Such reports suggest that the ‘sense of presence’ does not merely refer to the lack of cybersickness.

(c) The *something else* interpretation allows that the ‘sense of presence’ reported in immersive VR is an *additional* phenomenal feature that is not an aspect of non-immersive VR. However, it denies that this additional phenomenal feature is best explained by the hypothesis that immersive VR experiences have non-conceptual *de se* content. There is no doubt that some of the phenomenal differences between the overall experiences induced by immersive and non-immersive VR respectively are not simply due to the presence or absence of perceptual experience with *de se* content. For example, the overall experience of immersive VR typically involves sensations of bodily movement, since it allows users to move to control the visual viewpoint. By contrast, the overall experience of non-immersive VR might often lack such sensations, because users might stay still as they would when watching a movie on a flat screen. Consequently, at least in some cases, part of what ‘additional’ in the overall experience of immersive VR – compared to non-immersive VR – should involve sensations of bodily movement. However, this is not sufficient to explain the phenomenal contrast systematically reported between immersive and non-immersive VR. Indeed, users of non-immersive VR can and occasionally do move their heads (voluntarily or involuntarily), even though doing so has no effect on the graphical input to the head-mounted display because of the lack of head-tracking. If the phenomenal contrast between immersive and non-immersive VR was merely a matter of experiencing sensations of bodily movement, one would expect users who move their heads in non-immersive VR would report a sense of presence in the virtual environment; but this is not what the empirical evidence suggests. Consequently, there must be something additional to the experience

<sup>18</sup>[www.reddit.com/r/Vive/comments/8mxq15/vive\\_pro\\_elite\\_dangerous\\_first\\_vr\\_experience\\_to](http://www.reddit.com/r/Vive/comments/8mxq15/vive_pro_elite_dangerous_first_vr_experience_to).

of immersive VR, a sense of presence distinct from sensations of bodily movement. This sense of presence can plausibly be taken to reflect the non-conceptual *de se* content that perceptual experiences have in immersive VR, and lack in non-immersive VR.

In summary, objections to the three alternative interpretations of the phenomenal contrast between immersive and non-immersive VR support the hypothesis that perceptual experiences induced by immersive VR have non-conceptual *de se* content, while perceptual experiences induced by non-immersive VR have only *de hinc* content.

### 7.5 The agency-involving account revisited

I have argued so far that the phenomenal contrast between perspectively structured perceptual experiences induced by immersive VR and those induced by non-immersive VR is best explained by the hypothesis that the former, but not the latter, have non-conceptual *de se* content. This content represents the egocentric location of the point of origin of the perceptual experience's frame of reference as the subject's own location with respect to her environment. We can now connect this empirical hypothesis with a philosophical elucidatory account of what it takes for a perspectively structured perceptual experience to have non-conceptual *de se* content.

Recall that according to Peacocke (2017)'s *agency-involving account* of the subject's capacity to undergo experiences with non-conceptual *de se* content, there are two severally sufficient and jointly necessary conditions for a non-conceptual component *c* of intentional content employed by a creature to be the non-conceptual *de se* notion:

- (1) there is a range of action notions *A* for which the creature must be capable of being in mental states... with the content
 

*c* is *A*-ing

where the state... is produced by the initiation of an *A*-ing by the reference of *c*;  
and
- (2) there is a range of notions *F* of bodily properties, spatial properties, and past tense properties *F* such that the creature is capable of being in mental states... with the content *c* is *F*; where in these attributions,
 

*c* is *F*

is accepted (in central basic cases) if and only if

this body is *F*

is also accepted.

Peacocke (2017, p. 293)

Peacocke glosses these two conditions in simpler terms as follows:

[According to condition (1)] for possession of a first-person notion...the subject must be capable of representing the state of affairs that he or she or it is performing some action-type, such as that of moving, running, or extending an arm. [According to condition (2)] [T]he subject must also be capable of using in bodily attributions the same way of representing itself as is employed in those action-attributions.

Peacocke (2019, p. 122)

Note that in this explanation, Peacocke presents his account as an account of what is required for *possession* of the first-person notion, that is for the *capacity* to deploy the non-conceptual notion in the content of some mental state. Consequently, the *agency-involving account* does not explain why, for a subject who does possess the first-person notion, that notion might be deployed in some conscious mental states and not in others. Specifically, in the context of the hypothesis I have defended, I am interested in explaining why perspectively structured experiences induced by immersive VR would have non-conceptual *de se* content, while experiences induced by non-immersive VR would have only non-conceptual *de hinc* content – for the very same human subjects who possess the first-person notion.

The empirical literature on the sense of presence can help us answer this philosophical question. Remember that non-immersive VR differs from immersive VR in so far as non-immersive VR does not involve head-tracking capabilities. Consequently, the actual movements of the user's head in real three-dimensional space are not translated into movements of the point of origin of the virtual viewpoint within the virtual environment. This is true not only for *visual* input – i.e., the user's real head movements do not result in a shift of the visual scenery rendered in the display –, but also for *auditory* input – i.e., the user's real head movements do not result in a change in the timing and loudness of environmental sounds rendered in each of the user's ears through headphones. Thus, non-immersive VR does not provide *action-contingent visual and auditory feedback*, while immersive VR does. The fact that the latter, but not the former, induces a 'sense of presence' thus seems intimately connected to the availability of action-contingent feedback, as several VR psychologists have pointed out:

You know you are ‘there’ because sounds and images in the virtual world respond like the real world to your head movements.

Heeter (1992, p. 264)

The real-time update of sensory perception as a result of movement (e.g., head turning) gives rise to the sense of ‘being there’ – the illusory sensation of being in the computer-generated environment...

Slater and Sanchez-Vives (2016, p. 5)

Stereoscopic displays allow us to look at a scene from two different viewpoints... but [without head-tracking] this does not provide a sense of presence because the location does not behave like locations usually do – namely, changing as we move.

Troje (2019, p. 1036)

The association between the sense of presence and action-dependent *visual* feedback in particular is well-established. A large meta-analysis of the relationship between the sense of presence (as measured by various questionnaires) and different technical features related to the level of immersion of VR systems in eighty-eight studies found that head-tracking had the most significant impact on the occurrence of a sense of presence (Cummings and Bailenson 2016). In particular, tracking was found to be considerably more important than the resolution and visual quality of the rendered scene, suggesting that the sense of presence is not related to the degree of realism of the virtual environment. In other words, one does not feel as if one is present within a virtual environment simply because the virtual environment looks sufficiently similar to a real environment. Head-tracking was also found to have three times as much influence on self-reported presence as stereopsis (Snow and Williges 1998): while stereopsis mediates the perspectival spatial content of VR-experiences, the sense of presence appears to be specifically mediated by action-dependent feedback.

More precisely, head-tracking guarantees the congruency of (a) visual and auditory signals that provide information about changes in the egocentric locations of environmental landmarks (e.g., a seen tree moves from left to right across the subject’s visual field) with (b) vestibular and proprioceptive signals that provide information about the actual movement of the subject in three-dimensional space (e.g., the subject feels her head moving from right to left). In immersive VR, these signals are congruent and their integration provides the subject with a sense of where they are in their environment. However, when signals in category (a) are conflicting with signals of category (b), as in non-immersive VR, the

subject's perceptual experience can no longer represent the location of the point of origin of its egocentric spatial frame of reference as the subject's own location.

In accordance with this interpretation of the empirical evidence, I suggest that a perceptual experience *e* (e.g., a visual experience of a tree) of a subject *S* has non-conceptual *de se* content (e.g., <*i* am in front of a tree>, where *i* is the non-conceptual analogue of the first-person concept) if and only if:

- (1) experience *e* has perspectival spatial content (grounded in the availability of depth cues, particularly through stereopsis) in virtue of which it represents the location of a point of origin with respect to the environment (e.g., it represents the origin as being in front of the tree); and
- (2) The spatial information conveyed by the sensory signals (e.g., visual) whose processing experience *e* depends on are congruent with vestibular and proprioceptive signals about *S*'s bodily movement in three-dimensional space.

If both of these conditions are met, then *e* should not merely have *de hinc* content (e.g., <*here* is in front of a tree>), but have genuine *de se* content (e.g., <*i* am in front of a tree>), because it represents the location from which the subject perceives the environment as the location from which the subject can move herself in the environment. Note that this does not actually require the subject to move her body: if, say, visual input is fixed while the subject does not move, there is no conflict between visual cues on the one hand, and vestibular/proprioceptive cues on the other. Nonetheless, the congruence between the two types of signals is constantly put to the test, as it were, by small involuntary head movements known as 'head jitters'; for example, such movements have been shown to provide consistent and reliable cues about self-induced motion parallax in immersive VR (Fulvio and Rokers 2017).

This account of the *de se* content of perceptual experiences deserves to be called an 'agency-involving' account, because it highlights an important link between perception and action in self-representation. This is consistent with the fact that reports of the sense of presence in immersive VR are highly correlated with a number of behaviours in reaction to visual stimuli designed to elicit bodily movement. Thus, Nichols et al. (2000) presented

participants with a sudden looming stimulus towards their viewpoint in VR, and found that participants who scored higher on the SUS presence questionnaire exhibited significantly more startle responses to this stimulus. T. L. Y. Wu et al. (2019) found that the postural sway of participants riding a virtual rollercoaster in VR was significantly more pronounced in the head-tracking condition than in the passive conditions, and that this difference was positively correlated with the participants' scores on the IPQ questionnaire. The sense of presence was also found to be consistently correlated with the strength of the fear responses of phobic patients in VR exposure therapy (Alsina-Jurnet et al. 2011; Price et al. 2011; Riva et al. 2007; Robillard et al. 2003). These empirical findings are not surprising if the availability of action-contingent feedback mediates the non-conceptual *de se* content of perceptual experiences, allowing such experiences to represent the subject's location as the location from which she can act.

The upshot of this analysis is that ordinary visual and auditory experiences, in so far as they normally have perspectival spatial content and provide action-contingent feedback, normally have non-conceptual *de se* content rather than merely *de hinc* content. Consequently, in ordinary circumstances, a subject undergoing a conscious visual or auditory experience is *ipso facto* phenomenally self-conscious, without needing to engage in *de se* thinking. We can refer to this third kind of phenomenal self-consciousness as *spatial self-consciousness*.

## 7.6 Spatial self-consciousness in ordinary experience

In the previous chapter, I suggested that bodily self-consciousness is rather prevalent in ordinary experience, but probably not ubiquitous. What about spatial self-consciousness? It seems plausible that healthy human beings constantly undergo visual and/or auditory experiences in the ordinary wakeful state, from the moment they wake up to the moment they drift into sleep. Interestingly, Schwitzgebel (2007)'s experience sampling experiment also tested a group of participants on the frequency of their visual experiences; that is, after hearing a 'beep' at random intervals, participants had to report whether or not they had a visual experience just prior to the 'beep'. Schwitzgebel found that the median frequency

of reports of visual experiences across participants was 100%, supporting the intuition that visual experience is virtually ubiquitous in the wakeful state.

Given that the processing of egocentric distance and bearing through left-right disparities and other cues is not under voluntary control and built into the visual and auditory systems, it is also plausible that virtually all visual and auditory experiences we have in the wakeful state have *perspectival spatial content*. Furthermore, because of the basic anatomical fact that the eyes and ears are located on the subject's head, visual and auditory signals are normally narrowly coupled with vestibular and proprioceptive information about self-motion, thus providing congruent *action-contingent feedback*. On the account I have offered in this chapter, it follows first that visual and auditory experiences normally have non-conceptual *de se* content, and second that for healthy (sighted and hearing) subject, phenomenal self-consciousness is ubiquitous in the ordinary wakeful state.

Thus, the account of the non-conceptual *de se* content of perspectively structured perceptual experiences provided in this section appears to vindicate at least a restrictive version of (TSC) through the following argument:

- (P1) For any healthy neurotypical human subject S, if S is conscious in the ordinary wakeful state, then S undergoes visual and/or auditory experiences with non-conceptual *de se* content.
- (P2) A subject S undergoing visual and/or auditory experiences with non-conceptual *de se* content is *ipso facto* phenomenally self-conscious.
- (C1) Therefore, for any healthy neurotypical human subject S, if S is conscious in the ordinary wakeful state, then S is phenomenally self-conscious.

## 7.7 Conclusion

At the beginning of Part II, I indicated that my goal was to assess the empirical plausibility of the typicalist claim (TSC), according to which conscious subjects are phenomenally self-conscious *in ordinary circumstances*. In chapters 5 and 6, I argued that one can be phenomenally self-conscious by engaging in conscious *de se* thinking (*cognitive self-consciousness*) and by undergoing bodily experiences with non-conceptual *de se* content

(*bodily self-consciousness*). However, it is plausible that neither cognitive nor bodily self-consciousness are ubiquitous in the ordinary wakeful state. In this chapter, I argued that one can be phenomenally self-conscious by undergoing perceptual experiences with non-conceptual *de se* content (*spatial self-consciousness*). Since it is plausible that such experiences are virtually ubiquitous in the ordinary wakeful state, (TSC) is vindicated. This analysis leaves open several important questions about phenomenal self-consciousness, on which I will now offer some concluding remarks.

## Conclusion:

# The Significance of Self-Consciousness

I have stated in the introduction that the broad aim of this thesis was to illuminate the relationship between consciousness and self-consciousness. We are now in a good position to answer our original questions about this relationship: Does consciousness constitutively involve self-consciousness? Is self-consciousness an aspect of every conscious mental state? If not, does self-consciousness nonetheless occupy a particularly significant place in our conscious mental lives? It is impossible to address these questions without getting clearer on what self-consciousness might be. To this end, I have disambiguated the notion of self-consciousness at two levels. At the higher level, I have distinguished between two broad notions of self-consciousness:

- (1) the ‘non-egological’ notion of self-consciousness, or consciousness of *consciousness itself*, or consciousness of *one’s experience*; and
- (2) the ‘egological’ notion of self-consciousness, or consciousness of *oneself*.

At the lower level, each of these broad notions of self-consciousness can itself be further disambiguated into more specific notions that fall within two categories:

- (a) ‘thin’ or ‘minimal’ notions that point to foundational aspects of phenomenology, and in particular to the second-order notion of *phenomenality*: there being *something*, rather than *nothing*, that it is like for a subject S to be in a conscious mental state M (or, equivalently, S being in M’s making *some* constitutive contribution to S’s overall phenomenology, rather than *none*); and
- (b) ‘thick’ or ‘strong’ notions that point to further, more determinate, aspects of phenomenology, which are taken to make their own, distinct contribution to the subject’s overall phenomenology.

The notions subsumed under the terms ‘pre-reflective self-consciousness’ or ‘pre-reflective self-awareness’ in the phenomenological tradition fall within category (a), whether they are glossed as (1.a) pre-reflective consciousness or awareness *of one’s experience*, or (2.a) pre-reflective consciousness or awareness *of oneself*. Some of the notions subsumed under the terms ‘subjective character’ or ‘for-me-ness’ also fall within categories (1.a) and (2.a), but others – such as the notion Kriegel refers to as ‘inner awareness’ – fall within category (1.b). Finally, the notion that I have called ‘phenomenal self-consciousness’ falls within category (2.b): it refers to a consciousness *of oneself as oneself* that makes a more determinate contribution to the subject’s overall phenomenology.

In Part I, I argued against what I provocatively called the ‘myth’ of constitutive self-consciousness: the claim that self-consciousness is *constitutive of* – or a *necessary aspect of* – consciousness, and consequently an aspect of every conscious mental state. Naturally, the assessment of this claim depends on how one understands the relevant notion of self-consciousness.

This claim may well be *true* if the notion of self-consciousness that one deems constitutive of consciousness is a ‘thin’ or ‘minimal’ notion that falls under categories (1.a) or (2.a). However, the Nagelian account of consciousness provides us with a framework in which to elucidate the aspects of consciousness that such notions seek to illuminate without appealing to a term as polysemous as ‘self-consciousness’. Rather than illuminating the relevant aspects of consciousness, the use of the term ‘self-consciousness’, even when prefixed with a cautionary ‘pre-reflective’, runs the risk of leading readers to mistake the thin notion for a thicker one. Consequently, I suggested, if a notion of self-consciousness that *is* constitutive of consciousness is ‘thin’ – belonging to category (1.a) or (2.a) – then one would ideally avoid referring to the notion with the term ‘self-consciousness’.

The claim that self-consciousness is constitutive of consciousness might, on the other hand, involve a ‘thicker’ notion of self-consciousness – belonging to categories (1.b) or (2.b). I argued that such claims are not adequately supported. Contentious intuitions about phenomenology do not provide us with good enough reasons to believe that some constitutive aspect of consciousness makes its own, distinct, determinate contribution to the phenomenal character of every conscious mental state.

Dispelling the myth of constitutive self-consciousness allows us to examine the proper place of ‘thicker’ notions of self-consciousness in our conscious mental lives. Among these, the notion of phenomenal self-consciousness – in category (2.b) – is perhaps the most deserving of the name ‘self-consciousness’ given its connection to the representation of the self as such. In Part II, I took a closer look at the typicalist claim that consciousness comes along with phenomenal self-consciousness in ordinary circumstances, as a matter of contingent empirical fact:

(TSC) *In ordinary circumstances*, if a subject S is conscious, (a) S is conscious of S [SELF], and (b) S’s being conscious of S [SELF] makes a determinate constitutive contribution to S’s overall phenomenology at *t*.

According to (TSC), phenomenal self-consciousness is not a *constitutive* feature of consciousness; but it is nonetheless a *typical* feature of conscious experience. A number of philosophers seem to find (TSC) intuitively plausible, although difficult to defend because the relevant aspect of phenomenology seems “very hard to pin down” (Chalmers 1996, p. 10). My strategy in Part II has been to examine in more detail three candidate determinates of which phenomenal self-consciousness is a determinable:

- *Cognitive self-consciousness*. A subject S is cognitively self-conscious if and only if S engages in conscious thinking with conceptual *de se* content (i.e., S consciously thinks about herself as herself).
- *Bodily self-consciousness*. A subject S is bodily self-conscious if and only if S undergoes a bodily experience with non-conceptual *de se* content (i.e., S is conscious of herself as the bodily subject whose body part S feels a sensation in or on).
- *Spatial self-consciousness*. A subject S is spatially self-conscious if and only if S undergoes a perspectively structured perceptual experience with non-conceptual *de se* content (i.e., S is conscious of herself as being located at the origin of her perceptual frame of reference).

Rather than simply *assuming* that these three forms of phenomenal self-consciousness are present, let alone ubiquitous, in ordinary conscious experience, I have provided empirical arguments for their existence as genuine and distinct aspects of phenomenology, as well as elucidatory accounts of what it takes for conscious thoughts, bodily experiences and perspectively structured perceptual experiences to have (conceptual or non-conceptual) *de se* content.

Furthermore, I have assessed the prevalence of each of these forms of phenomenal self-consciousness in conscious experience. I have argued that each of these can be missing in some cases: a subject can be conscious without engaging in conscious *de se* thinking, or without undergoing a bodily experience with *de se* content, or without undergoing a perspectively structured perceptual experiences with *de se* content. This is an *exclusive* disjunction. I have not (yet) argued that a subject can be conscious while cognitive, bodily and spatial self-consciousness are all absent.

The order in which I have considered these three forms of phenomenal self-consciousness matches their order from the least pervasive to the most pervasive in the ordinary wakeful state. We often think about ourselves as ourselves; but even the most narcissistic individuals do not go through life with incessant *de se* thoughts. More frequently still, we undergo bodily experiences with *de se* content. While the saliency of such experiences certainly waxes and wanes depending on how much we attend to bodily sensations, it is plausible that we very often experience *some* bodily sensation or other in *some* body part or other, and that, in ordinary circumstances, such experience has non-conceptual *de se* content. Finally, visual and auditory experiences that have not only perspectival spatial content but also non-conceptual *de se* content seem virtually ubiquitous in ordinary experience. There is hardly a moment in our conscious waking lives when we do not hear sounds or see our environment from a perspective, in such a way that we are conscious of ourselves as being located at the origin of our perceptual frame of reference.

The ubiquity of spatial self-consciousness alone appears to support a version of (TSC) in which the relevant set of 'ordinary circumstances' is (conservatively) restricted to the conscious wakeful state of healthy, neurotypical adult human beings. *A fortiori*, the account I have given of all three forms of phenomenal self-consciousness provides even stronger

support for the following disjunctive claim – where the disjunction is *inclusive* – that entails the restrictive interpretation of (TSC):

(DIS) For any healthy neurotypical adult human subject S, if S is conscious in the ordinary wakeful state, then S is cognitively self-conscious, or bodily self-conscious, or spatially self-conscious.

It follows from (DIS) that in the ordinary conscious wakeful state, healthy neurotypical adult human subjects are phenomenally self-conscious. There is indeed a ‘sense of self’ in ordinary conscious experience, but it is not detached, as it were, from cognitive, bodily and perceptual mental states; it is not a free-floating ‘self *quale*’. Ordinary forms of self-consciousness are genuine aspects of ordinary phenomenology, but this is because they are aspects of the phenomenology of ordinary experiences – thinking thoughts, feeling bodily sensations, seeing and hearing our environment. Hume was right to point out that one “can never catch [one’s] self at any time without a perception” (Hume 1978, p. 252), or, more exactly, without a particular conscious experience that represents the world and oneself as being a certain way, be it a conscious thought, a bodily experience, or a perception.

Many substantive and philosophically interesting questions remain about phenomenal self-consciousness, and these questions could occupy many more chapters or, indeed, another dissertation. Although most of these questions will have to be addressed in future work, I will briefly consider three of them to conclude this investigation. The first question is an obvious one: while I have limited my analysis to three forms of phenomenal self-consciousness, might there be other forms of self-consciousness that are not reducible to cognitive, bodily and spatial self-consciousness? I am inclined to answer this question in the affirmative. For example, many authors believe that there is such a thing as a sense of agency that normally accompanies voluntary movement (e.g., Bayne 2008; Haggard 2017), and perhaps even mental actions (e.g., Peacocke 2006; Proust 2009). This sense of agency is typically defined as a form of consciousness of one’s actions as one’s own, or, more appropriately, a consciousness of oneself as the agent of one’s actions. Under such a definition, the sense of agency would be yet another determinate of phenomenal self-consciousness – presumably not reducible to the other three. The existence of a phenomenology of agency

can also be defended through arguments from phenomenal contrast; it would explain, for example, the apparent contrast between feeling an involuntary muscle twitch in a body part, and twitching the same muscle voluntarily. While, for lack of space, I did not examine such phenomenology, I find it intuitively plausible that it does occur in ordinary experience, and quite frequently (given the frequency of voluntary movements).

Another candidate determinate of phenomenal self-consciousness that I have left aside is so-called 'self-conscious emotion' (see Sznycer 2019). Indeed, conscious emotions such as pride and shame often seem to involve some form of self-representation: one can feel proud or ashamed *of oneself*. The nature of emotional phenomenology is hotly debated.<sup>19</sup> It is possible, for example, that 'self-conscious emotions' like pride and shame are reducible to a combination of some bodily feeling (e.g., elation or anxiety) with a conscious *de se* thought. In this case, self-conscious emotion would not constitute a genuinely distinct form of phenomenal self-consciousness. I do not find this reductionist suggestion particularly implausible, but the question would have to be settled by a substantial discussion of the phenomenology of emotion. In any case, it seems to me that self-conscious emotions are rather sporadic occurrences in our mental lives.

Are there other forms of phenomenal self-consciousness that are not aspects of cognitive, bodily, perceptual, agentive or emotional phenomenology? Perhaps, in so far as there is a *sui generis* phenomenology of imagination, there is a form of imaginative self-consciousness that consists in imagining oneself as such in some situation.<sup>20</sup> I do not have strong intuitions about this suggestion; but as in the case of self-conscious emotions, imaginative self-consciousness – if there is such a thing – should be rather infrequent.

This brings us to the second outstanding question: are there states of consciousness in which a subject can lack *any* form of phenomenal self-consciousness? I have argued that self-consciousness is not *constitutive* of consciousness; but it could still be the case, as a matter of contingent empirical fact, or perhaps even as a matter of nomological necessity for human beings, that we are phenomenally self-conscious whenever we are conscious at

<sup>19</sup>See Kriegel (2015, chapter 4) for a review and discussion.

<sup>20</sup>See Kriegel (2015, conclusion, §1) for a defence of the claim that there is a *sui generis* phenomenology of imagination.

all. Part of the difficulty of answering this question is that it would require us to have an *exhaustive* inventory of the different ways in which one can be phenomenally self-conscious, in order to determine whether they are *all* missing from some particular conscious mental state(s). This project is deemed to be controversial, because for any putative example of a selfless state of consciousness, someone is likely to suggest that it does instantiate some fundamental and elusive form of phenomenal self-consciousness that one has failed to account for. Perhaps this kind of dialectic is part of the reason why some philosophers do not draw a sharp distinction between consciousness and self-consciousness:

[W]e often fail to distinguish clearly enough between consciousness and self-consciousness. One possible explanation of this failure is that we have no access to unproblematic actual examples of consciousness without self-consciousness.

S. L. White (1987, p. 168)

Putative examples of consciousness without (phenomenal) self-consciousness will probably never be ‘unproblematic’, if by this we simply mean ‘uncontroversial’. But I do think such examples exist, and are of tremendous interest for philosophy of mind. In particular, I have argued elsewhere that there is a subset of drug-induced states (known in the empirical literature as ‘drug-induced ego dissolution’), as well as a subset of meditation-induced states, that can plausibly be taken to lack any form of phenomenal self-consciousness – any form of conscious self-representation.<sup>21</sup> Subjects in these states are effectively real-world examples of what Peacocke (2014) calls a creature at ‘degree 0’ of self-representation. Studying the phenomenology and neurophysiology of drug-induced and meditation-induced loss of self-consciousness is perhaps the best antidote to what Dennett (1991) calls Philosopher’s Syndrome – “mistaking a failure of imagination for an insight into necessity” (p. 401) – when it comes to phenomenal self-consciousness. It is perhaps difficult to imagine what it would be like to lack phenomenal self-consciousness, because self-consciousness is so pervasive in the wakeful state. But it does not follow that self-consciousness cannot actually be completely missing from conscious experience.

The third and final question concerns the significance of phenomenal self-consciousness, and specifically of the three forms of self-consciousness I have chosen to

<sup>21</sup>See Millière (forthcoming; 2017) and Millière et al. (2018) for a discussion.

focus on in the present work. The significance of cognitive self-consciousness is fairly obvious: it is the only way in which we can consciously represent ourselves to ourselves by means of the first-person concept. As such, it is directly connected to our beliefs about ourselves (*de se* beliefs) and thus to self-knowledge. Standing *de se* beliefs are dispositional states, but when we bring the contents of such beliefs to consciousness, we do so by engaging in conscious *de se* thinking. Cognitive self-consciousness is a central component of what it is to reflect upon what we believe, and on what we know, about ourselves.

Bodily and spatial self-consciousness are also connected to *de se* belief and self-knowledge, but as *sources* rather than *outputs* of *de se* belief and knowledge. Undergoing a bodily experience or a perspectively structured perceptual experience with non-conceptual *de se* content puts us in a position rationally to *judge* and *believe* that we have certain physical and spatial properties, and most of the time to *know* that we have such properties. In normal circumstances, feeling a pressure on one's thigh is sufficient to justify the belief that one is being touched on one's thigh; likewise, seeing a tree in front of oneself is sufficient to justify the belief that one is front of a tree.

The role of bodily and spatial self-consciousness in certain kinds of *de se* beliefs and self-knowledge relates to a further point about their significance. Being conscious of oneself as a bodily subject located within one's environment is presumably an important aspect of the survival of conscious organisms. As de Vignemont (2018, p. 176) emphasises, our bodies are not like any other objects in the world, they need to be protected from threats if we are to survive. Bodily self-consciousness enables us to feel bodily sensations not simply as any other event in the world, but as events that happen *to ourselves* as bodily subjects. Likewise, spatial self-consciousness enables us to act on and react to our environment; this is why subjects in non-immersive VR who do not have a 'sense of presence' fail to react to looming or arousing stimuli. In sum, if a creature is to consciously navigate its environment and avoid predators, it seems important that it be conscious of its body as its own (rather than any other part of the environment) and of the origin of its perceptual frame of reference as its own location (rather than a location that has no relevance to its actions).

The upshot of this discussion is that phenomenal self-consciousness is not only a pervasive aspect of our waking lives, but also plays a significant role in a variety of phenomena. In

cognitively sophisticated creatures such as human beings, it grounds beliefs about ourselves and is part of what it is to reflect upon these beliefs; and in conscious creatures more generally, it plays an important role in guiding behaviour for survival. Despite its pervasiveness and significance, however, phenomenal self-consciousness should not be confused with consciousness: being phenomenally self-conscious is *ipso facto* being conscious, but being conscious is not *ipso facto* being phenomenally self-conscious.

## References

- Aimola Davies, Anne M. and Rebekah C. White (2013). “A Sensational Illusion: Vision-Touch Synaesthesia and the Rubber Hand Paradigm”. In: *Cortex* 49.3, pp. 806–818.
- Alsina-Jurnet, Ivan, José Gutiérrez-Maldonado, and María-Virgina Rangel-Gómez (2011). “The Role of Presence in the Level of Anxiety Experienced in Clinical Virtual Environments”. In: *Computers in Human Behavior* 27.1, pp. 504–512.
- Alsmith, Adrian (2015). “Mental Activity and the Sense of Ownership”. In: *Review of Philosophy and Psychology* 6.4, pp. 881–896.
- (2017). “Perspectival Structure and Agentive Self-Location”. In: *The Subject’s Matter: Self-Consciousness and the Body*. Ed. by Adrian Alsmith and Frédérique de Vignemont. Oxford: Oxford University Press, p. 263.
- Andrews-Hanna, Jessica R. et al. (2013). “A Penny for Your Thoughts: Dimensions of Self-Generated Thought Content and Relationships with Individual Differences in Emotional Wellbeing”. In: *Frontiers in Psychology* 4.
- Armstrong, David M. (1962). *Bodily Sensations*. London and New York: Routledge.
- (1968). *A Materialist Theory of the Mind*. London and New York: Routledge.
- Arnauld, Antoine (1660/1775). “Lettre à La Soeur Marguerite Gertrude”. In: *Lettres de Messire Antoine Arnauld*. Vol. 1. Paris: Sigismond d’Arnay et Compagnie.
- (1683/2011). *Des vraies et des fausses idées*. Trans. by Denis Moreau. Paris: Vrin.
- Arnold, Dan (2010). “Self-Awareness (Svasamvitti) and Related Doctrines of Buddhists Following Dignāga: Philosophical Characterizations of Some of the Main Issues”. In: *Journal of Indian Philosophy* 38.3, pp. 323–378.
- Augustine (2002). *On the Trinity Books 8-15*. Ed. by Gareth B. Matthews. Trans. by Stephen McKenna. Cambridge and New York: Cambridge University Press.
- Bain, A. (1894). “Definition and Problems of Consciousness”. In: *Mind* 3.11, pp. 348–361.
- Baird, Benjamin, Jonathan Smallwood, and Jonathan W. Schooler (2011). “Back to the Future: Autobiographical Planning and the Functionality of Mind-Wandering”. In: *Consciousness and Cognition*. From Dreams to Psychosis: A European Science Foundation Exploratory Workshop 20.4, pp. 1604–1611.
- Baker, Lynne Rudder (2000). *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.
- Barnes, Jennifer (2001). “The Lived Experience of Meditation”. In: *Indo-Pacific Journal of Phenomenology* 1.2.
- Bayne, Tim (2004). “Self-Consciousness and the Unity of Consciousness”. In: *The Monist* 87.2, pp. 219–236.
- (2008). “The Phenomenology of Agency”. In: *Philosophy Compass* 3.1, pp. 182–202.
- Bayne, Tim and Maja Spener (2010). “Introspective Humility”. In: *Philosophical Issues* 20.1, pp. 1–22.
- Berker, Selim (2008). “Luminosity Regained”. In: *Philosophers’ Imprint* 8, pp. 1–22.
- Bermudez, Jose Luis (1998). *The Paradox of Self-Consciousness*. Cambridge, MA: MIT Press.

- Bermúdez, José Luis (2005). “The Phenomenology of Bodily Awareness”. In: *Phenomenology and Philosophy of Mind*. Ed. by David Woodruff Smith and Amie Lynn Thomasson. Oxford: Clarendon Press.
- (2011). “Bodily Awareness and Self-Consciousness”. In: *The Oxford Handbook of the Self*. Ed. by Shaun Gallagher. Oxford: Oxford University Press.
- (2017). “Ownership and the Space of the Body”. In: *The Subject’s Matter: Self-Consciousness and the Body*. Ed. by Adrian Alsmith and Frédérique de Vignemont. Cambridge, MA: MIT Press, pp. 117–144.
- (2018). “Bodily Ownership, Psychological Ownership, and Psychopathology”. In: *Review of Philosophy and Psychology*.
- Bicanski, A. and N. Burgess (2018). “A Neural-Level Model of Spatial Memory and Imagery.” In: *eLife* 7.
- Billon, Alexandre (2017). “Mineness First: Three Challenges to the Recent Theories of the Sense of Bodily Ownership”. In: *The Subject’s Matter: Self-Consciousness and the Body*. Ed. by Frédérique de Vignemont and Adrian Alsmith. Cambridge, MA: MIT Press, pp. 189–216.
- Bitbol, Michel and Claire Petitmengin (2017). “Neurophenomenology and the Micro-Phenomenological Interview”. In: *The Blackwell Companion to Consciousness*. Ed. by Susan Schneider and x Velmans. John Wiley & Sons, Ltd, pp. 726–739.
- Block, Ned (1995). “On a Confusion About a Function of Consciousness”. In: *Brain and Behavioral Sciences* 18.2, pp. 227–247.
- Blumenthal, James (2004). *Ornament of the Middle Way: A Study of the Madhyamaka Thought of Shantarakshita*. 1 edition. Ithaca, N.Y: Snow Lion Publications.
- Bottini, Gabriella et al. (2002). “Feeling Touches in Someone Else’s Hand”. In: *NeuroReport* 13.2, p. 249.
- Bourget, David (2017). “Why Are Some Phenomenal Experiences ‘Vivid’ and Others ‘Faint’? Representationalism, Imagery, and Cognitive Phenomenology”. In: *Australasian Journal of Philosophy* 95.4, pp. 673–687.
- Boyle, Matthew (2011). “Transparent Self-Knowledge”. In: *Aristotelian Society Supplementary Volume* 85.1, pp. 223–241.
- Braddon-Mitchell, David and Frank Jackson (2007). *The Philosophy of Mind and Cognition*. Malden, MA: Blackwell.
- Bradley, Adam (2019). “The Feeling of Bodily Ownership”. In: *Philosophy and Phenomenological Research*.
- Brentano, Franz (1874/1995). *Psychology from an Empirical Standpoint*. Ed. by Oskar Kraus and Linda L. McAlister. Paperback ed. International Library of Philosophy. London ; New York: Routledge.
- Byrne, Patrick, Suzanna Becker, and Neil Burgess (2007). “Remembering the Past and Imagining the Future”. In: *Psychological review* 114.2, pp. 340–375.
- Calkins, Mary Whiton (1908). “Psychology as Science of Self: II. The Nature of the Self”. In: *The Journal of Philosophy, Psychology and Scientific Methods* 5.3, pp. 64–68.
- Campbell, J. (1994). *Past, Space, and Self*. Cambridge, MA: MIT Press.
- Carruthers, Peter (2004). “Hop Over FOR, HOT Theory”. In: *Higher-Order Theories of Consciousness: An Anthology*. Ed. by Rocco J. Gennaro. John Benjamins.
- (2005). “Conscious Experience Versus Conscious Thought”. In: *Consciousness and Self-Reference*. Ed. by Uriah Kriegel and Kenneth Williford. Cambridge, MA: MIT Press.
- Carruthers, Peter and Bénédicte Veillet (2011). “The Case Against Cognitive Phenomenology”. In: *Cognitive Phenomenology*. Ed. by Tim Bayne and Michelle Montague. Oxford: Oxford University Press, p. 35.

- Cassam, Quassim (1997). *Self and World*. Oxford: Oxford University Press.
- Caston, Victor (2002). "Aristotle on Consciousness". In: *Mind* 111.444, pp. 751–815.
- Chalmers, David J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Child, William (2011). "Wittgenstein on the First Person". In: *The Oxford Handbook of Wittgenstein*. Ed. by Marie McGinn and Oskari Kuusela. Oxford: Oxford University Press, pp. 375–402.
- Christoff, Kalina et al. (2009). "Experience Sampling during fMRI Reveals Default Network and Executive System Contributions to Mind Wandering". In: *Proceedings of the National Academy of Sciences* 106.21, pp. 8719–8724.
- Chudnoff, Elijah (2015a). *Cognitive Phenomenology*. New York and London: Routledge.
- (2015b). "Phenomenal Contrast Arguments for Cognitive Phenomenology". In: *Philosophy and Phenomenological Research* 91.1, pp. 82–104.
- Cobb, Sue V. G. et al. (1999). "Virtual Reality-Induced Symptoms and Effects (VRISE)". In: *Presence: Teleoperators and Virtual Environments* 8.2, pp. 169–186.
- Coltheart, Max, Robyn Langdon, and Ryan McKay (2011). "Delusional Belief". In: *Annual Review of Psychology* 62.1, pp. 271–298.
- Coltheart, Max, Peter Menzies, and John Sutton (2010). "Abductive Inference and Delusional Belief". In: *Cognitive Neuropsychiatry* 15.1-3, pp. 261–287.
- Coseru, Christian (2016). "Dignaga and Dharmakirti on Perception and Self-Awareness". In: *The Buddhist World*. Ed. by John Powers. London and New York: Routledge, pp. 526–537.
- Coventry, Angela and Uriah Kriegel (2008). "Locke on Consciousness". In: *History of Philosophy Quarterly* 25.3, pp. 221–242.
- Crane, Tim (1999). "The Mind-Body Problem". In: *The MIT Encyclopedia of the Cognitive Sciences*. Ed. by Rob Wilson and Frank Keil. Cambridge, MA: MIT Press.
- (2013). "Unconscious Belief and Conscious Thought". In: *Phenomenal Intentionality*. Ed. by Uriah Kriegel. Oxford: Oxford University Press, pp. 156–173.
- Csikszentmihalyi, Mihaly (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper Collins.
- Cummings, James J. and Jeremy N. Bailenson (2016). "How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence". In: *Media Psychology* 19.2, pp. 272–309.
- D'Argembeau, Arnaud (2016). "The Role of Personal Goals in Future-Oriented Mental Time Travel". In: *Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel*. Ed. by Kourken Michaelian, Stanley B. Klein, and Karl K. Szpunar. Oxford: Oxford University Press, pp. 199–216.
- (2018). "Mind-Wandering and Self-Referential Thought". In: *The Oxford Handbook of Spontaneous Thought*. Ed. by Kieran C. R. Fox and Kalina Christoff. Oxford: Oxford University Press.
- Dainton, Barry (2008). *The Phenomenal Self*. Oxford: Oxford University Press.
- Damasio, Antonio (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. San Diego, CA: Houghton Mifflin Harcourt.
- Davies, Martin et al. (2001). "Monothematic Delusions: Towards a Two-Factor Account". In: *Philosophy, Psychiatry, & Psychology* 8.2, pp. 133–158.
- de Vignemont, Frédérique (forthcoming). "The Phenomenology of Bodily Ownership". In: *The Sense of Mineness*. Ed. by M. Garcia-Carpintero and M. Guillot. Oxford: Oxford University Press.
- (2007). "Habeas Corpus: The Sense of Ownership of One's Own Body". In: *Mind & Language* 22.4, pp. 427–449.

- de Vignemont, Frédérique (2013). "The Mark of Bodily Ownership". In: *Analysis* 73.4, pp. 643–651.
- (2018). *Mind the Body: An Exploration of Bodily Self-Awareness*. Oxford: Oxford University Press.
- (2019a). "Mind the Body (2) A Phenomenal Contrast for Bodily Ownership". In: *The Brains Blog*. URL: <http://philosophyofbrains.com/2019/02/26/mind-the-body-2-a-phenomenal-contrast-for-bodily-ownership.aspx> (visited on 01/11/2020).
- (2019b). "Mind the Body (4) What Kind of First-Personal Content?" In: *The Brains Blog*. URL: <http://philosophyofbrains.com/2019/02/28/mind-the-body-4-what-kind-of-first-personal-content.aspx> (visited on 01/12/2020).
- Dennett, Daniel C. (1991). *Consciousness Explained*. London: Penguin Books.
- DeRose, Keith (2002). "Review of T. Williamson, Knowledge and Its Limits". In: *The British Journal for the Philosophy of Science* 53.4, pp. 573–577.
- Descartes, René (1644/1983). *Principles of Philosophy*. Trans. by Valentine Rodger Miller and Reese P. Miller. Dordrecht: Reidel.
- (1648/1991). "Conversation with Burman". In: *The Philosophical Writings of Descartes*. Trans. by John Cottingham et al. Vol. 3. Cambridge: Cambridge University Press, pp. 332–353.
- (1641/2008). *Meditations on First Philosophy with Selections from the Objections and Replies*. Trans. by Michael Moriarty. Oxford: Oxford University Press.
- Dignāga (1968). *On Perception*. Trans. by Masaaki Hattori. Cambridge, MA: Harvard University Press.
- Dretske, Fred (1993). "Conscious Experience". In: *Mind* 102.406, pp. 263–283.
- Duncan, Matt (2018). "Subjectivity as Self-Acquaintance". In: *Journal of Consciousness Studies* 25.3-4, pp. 88–111.
- (2019). "The Self Shows Up in Experience". In: *Review of Philosophy and Psychology* 10.2, pp. 299–318.
- Dyck, Corey W. (2014). *Kant and Rational Psychology*. Oxford: Oxford University Press.
- Evans, Gareth (1982). *The Varieties of Reference*. Oxford: Oxford University Press.
- Farrell, B. A. (1950). "Experience". In: *Mind* 59.234, pp. 170–198.
- Farrell, J. (2016). "'What It Is Like' Talk Is Not Technical Talk". In: *Journal of Consciousness Studies* 23.9-10, pp. 50–65.
- Fasching, Wolfgang (2008). "Consciousness, Self-Consciousness, and Meditation". In: *Phenomenology and the Cognitive Sciences* 7.4, pp. 463–483.
- Finnigan, Bronwyn (2018). "Is Consciousness Reflexively Self-Aware? A Buddhist Analysis". In: *Ratio* 31.4, pp. 389–401.
- Flanagan, Owen J. (1992). *Consciousness Reconsidered*. Cambridge, MA: MIT Press.
- Frank, Manfred (2007). "Non-Objectal Subjectivity". In: *Journal of Consciousness Studies* 14.5-6, pp. 152–173.
- Frankfurt, Harry G. (1988). *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press.
- Frankish, K. (2016). "Not Disillusioned: Reply to Commentators". In: *Journal of Consciousness Studies* 23.11-12, pp. 256–289.
- Frege, Gottlob (1921/1956). "The Thought: A Logical Inquiry". Trans. by A. M. Quinton and Marcelle Quinton. In: *Mind* 65.259, pp. 289–311.
- Freud, Sigmund (1930/1961). *Civilization and Its Discontents*. Trans. by James Strachey. New York and London: W. W. Norton & Company.
- Friston, Karl (2018). "Am I Self-Conscious? (Or Does Self-Organization Entail Self-Consciousness?)" In: *Frontiers in Psychology* 9.

- Fulvio, Jacqueline M. and Bas Rokers (2017). "Use of Cues in Virtual Reality Depends on Visual Feedback". In: *Scientific Reports* 7.1, pp. 1–13.
- Gallagher, Maria and Elisa Raffaella Ferrè (2018). "Cybersickness: A Multisensory Integration Perspective". In: *Multisensory Research* 31.7, pp. 645–674.
- Gallagher, Shaun (2005). *How the Body Shapes the Mind*. Oxford: Oxford University Press.
- (2010). "Defining Consciousness: The Importance of Non-Reflective Self-Awareness". In: *Pragmatics and Cognition* 18.3, pp. 561–569.
- Gallagher, Shaun and Dan Zahavi (2012). *The Phenomenological Mind*. London and New York: Routledge.
- Gandola, Martina et al. (2012). "An Anatomical Account of Somatoparaphrenia". In: *Cortex* 48.9, pp. 1165–1178.
- Ganeri, Jonardon (2012). *The Self: Naturalism, Consciousness, and the First-Person Stance*. Oxford: Oxford University Press.
- Garfield, Jay L. (2015). *Engaging Buddhism: Why It Matters to Philosophy*. Oxford and New York: Oxford University Press.
- Gennaro, Rocco J. (2008). "Representationalism, Peripheral Awareness, and the Transparency of Experience". In: *Philosophical Studies* 139.1, pp. 39–56.
- Gertler, Brie (2012). "Conscious States as Objects of Awareness: On Uriah Kriegel, Subjective Consciousness: A Self-Representational Theory". In: *Philosophical Studies* 159.3, pp. 447–455.
- Gibson, James J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin.
- Giustina, Anna and Uriah Kriegel (2017). "Fact-Introspection, Thing-Introspection, and Inner Awareness". In: *Review of Philosophy and Psychology* 8.1, pp. 143–164.
- Goldman, Alvin I. (1970). *A Theory of Human Action*. Princeton, N.J: Princeton University Press.
- Gooch, Judith L., Monte H. Moore, and David K. Ryser (1993). "Prolonged Paralysis after Neuromuscular Junction Blockade: Case Reports and Electrodiagnostic Findings". In: *Archives of Physical Medicine and Rehabilitation* 74.9, pp. 1007–1011.
- Guillot, Marie (2017). "I Me Mine: On a Confusion Concerning the Subjective Character of Experience". In: *Review of Philosophy and Psychology* 8.1, pp. 23–53.
- Gurwitsch, Aron (1941). "A Non-Egological Conception of Consciousness". In: *Philosophy and Phenomenological Research* 1.3, pp. 325–338.
- (1953/2010). *The Field of Consciousness: Phenomenology of Theme, Thematic Field, and Marginal Consciousness*. Ed. by Richard M. Zaner and Lester E. Embree. Collected Works of Aron Gurwitsch (1901-1973) v. 3. Dordrecht ; New York: Springer.
- Hacker, P. M. S. (2002). "Is There Anything It Is Like to Be a Bat?" In: *Philosophy* 77.2, pp. 157–174.
- Hafting, Torkel et al. (2005). "Microstructure of a Spatial Map in the Entorhinal Cortex". In: *Nature* 436.7052, pp. 801–806.
- Haggard, Patrick (2017). "Sense of Agency in the Human Brain". In: *Nature Reviews Neuroscience* 18.4, p. 196.
- Halligan, Peter W., John C. Marshall, and Derick T. Wade (1995). "Unilateral Somatoparaphrenia After Right Hemisphere Stroke: A Case Description". In: *Cortex* 31.1, pp. 173–182.
- Harman, Gilbert (1990). "The Intrinsic Quality of Experience". In: *Philosophical Perspectives* 4.n/a, pp. 31–52.
- Hasenkamp, Wendy (2018). "Catching the Wandering Mind: Meditation as a Window into Spontaneous Thought". In: *The Oxford Handbook of Spontaneous Thought*. Ed. by Kieran C. R. Fox and Kalina Christoff. Oxford: Oxford University Press, pp. 539–553.
- Hasenkamp, Wendy et al. (2012). "Mind Wandering and Attention during Focused Meditation: A Fine-Grained Temporal Analysis of Fluctuating Cognitive States". In: *NeuroImage. Neuroergonomics: The Human Brain in Action and at Work* 59.1, pp. 750–760.

- Heal, Jane (2002). "On First-Person Authority". In: *Proceedings of the Aristotelian Society* 102.1, pp. 1–19.
- Hécaen, H. and J. de Ajariaguerra (1952). *Méconnaissances et Hallucinations Corporelles. Intégration et Désintégration de La Somatognosie*. Paris: Masson.
- Heeter, Carrie (1992). "Being There: The Subjective Experience of Presence". In: *Presence: Teleoperators and Virtual Environments* 1.2, pp. 262–271.
- Held, Richard M. and Nathaniel I. Durlach (1992). "Telepresence". In: *Presence: Teleoperators and Virtual Environments* 1.1, pp. 109–112.
- Hernández, Sergio Elías et al. (2016). "Increased Grey Matter Associated with Long-Term Sahaja Yoga Meditation: A Voxel-Based Morphometry Study". In: *PLOS ONE* 11.3, e0150757.
- Hinman, James R., G. William Chapman, and Michael E. Hasselmo (2019). "Neuronal Representation of Environmental Boundaries in Egocentric Coordinates". In: *Nature Communications* 10.1, pp. 1–8.
- Hopkins, G. M. (1959). *The Sermons and Devotional Writings of Gerard Manley Hopkins*. Ed. by Christopher Devlin. Oxford: Clarendon Press.
- Horgan, Terence E., John L. Tienson, and George Graham (2004). "Phenomenal Intentionality and the Brain in a Vat". In: *The Externalist Challenge*. Ed. by Richard Schantz. De Gruyter.
- Horgan, Terence and John Tienson (2002). "The Intentionality of Phenomenology and the Phenomenology of Intentionality". In: *Philosophy of Mind: Classical and Contemporary Readings*. Ed. by David J. Chalmers. Oxford and New York: Oxford University Press, pp. 520–533.
- Howell, Robert J. (2006). "Self-Knowledge and Self-Reference". In: *Philosophy and Phenomenological Research* 72.1, pp. 44–70.
- Howell, Robert J. and Brad Thompson (2017). "Phenomenally Mine: In Search of the Subjective Character of Consciousness". In: *Review of Philosophy and Psychology* 8.1, pp. 103–127.
- Hume, David (1978). *A Treatise of Human Nature*. Ed. by L. A. Selby-Bigge and P. H. Nidditch. 2nd edition. Oxford: Oxford University Press.
- Hurlburt, Russell T. and Sarah A. Akhter (2006). "The Descriptive Experience Sampling Method". In: *Phenomenology and the Cognitive Sciences* 5.3, pp. 271–301.
- Husserl, Edmund (1924/1959). *Erste Philosophie, 1923/24*. Ed. by Rudolf Boehm. Vol. 2. Martinus Nijhoff.
- (1931/1960). *Cartesian Meditations: An Introduction to Phenomenology*. Trans. by Dorion Cairns. The Hague: Nijhoff.
- (1920/1970). "Bewusstsein und Sinn — Sinn und Noema". In: *Analysen zur Passiven Synthesis: Aus Vorlesungs- und Forschungsmanuskripten 1918–1926*. Ed. by M. Fleischer. Den Haag: Springer, pp. 304–335.
- (1932/1973a). "Universale Geisteswissenschaft als Anthropologie. Sinn einer Anthropologie". In: *Zur Phänomenologie der Intersubjektivität: Texte aus dem Nachlass. Dritter Teil: 1929–1935*. Ed. by Ido Kern. Husserliana: Edmund Husserl – Gesammelte Werke. The Hague: Springer Netherlands, pp. 480–507.
- (1922/1973b). *Zur Phänomenologie der Intersubjektivität: Texte aus dem Nachlass. Zweiter Teil: 1921–1928*. Ed. by Kern. One thousand, nine hundred seventy-third. Dordrecht: Springer.
- (1907/1991a). "On the Dissolution of the Schema: Apprehension-Content-Apprehension". In: *On the Phenomenology of the Consciousness of Internal Time*. Trans. by John Barnett Brough. Dordrecht and Boston: Springer, pp. 279–346.
- (1905/1991b). "The Lectures on the Consciousness of Internal Time from the Year 1905". In: *On the Phenomenology of the Consciousness of Internal Time*. Trans. by John Barnett Brough. Dordrecht and Boston: Springer, pp. 3–103.

- Husserl, Edmund (1918/2001a). *Die Bernauer Manuskripte Über das Zeitbewusstsein (1917/18)*. Ed. by Rudolf Bernet and Dieter Lohmar. Husserliana: Edmund Husserl – Gesammelte Werke XXXIII. Springer Netherlands.
- (1901/2001b). *Logical Investigations*. Vol. 1. London and New York: Routledge.
- Invernizzi, Paola et al. (2013). “What Is Mine? Behavioral and Anatomical Dissociations between Somatoparaphrenia and Anosognosia for Hemiplegia”. In: *Behavioural Neurology* 26.1-2, pp. 139–150.
- Jackson, S.A. and R.C. Eklund (2004). *The Flow Scales Manual*. Morgantown, WV: Fitness Information Technology.
- Jackson, Susan A. and Herbert W. Marsh (1996). “Development and Validation of a Scale to Measure Optimal Experience: The Flow State Scale”. In: *Journal of Sport and Exercise Psychology* 18.1, pp. 17–35.
- James, William (1890). *The Principles of Psychology*. New York: Henry Holt and Company.
- (1892/1961). *Psychology: The Briefer Course*. Ed. by Gordon Allport. New York: Harper and Row.
- Janzen, Greg (2006). “Phenomenal Character as Implicit Self-Awareness”. In: *Journal of Consciousness Studies* 13.12, pp. 44–73.
- (2008). *The Reflexive Nature of Consciousness*. Vol. 72. Advances in Consciousness Research. Amsterdam: John Benjamins Publishing Company.
- Kahan, Tracey L. and Stephanie Claudatos (2016). “Phenomenological Features of Dreams: Results from Dream Log Studies Using the Subjective Experiences Rating Scale (SERS)”. In: *Consciousness and Cognition* 41, pp. 159–176.
- Kant, Immanuel (1781/1999). *Critique of Pure Reason*. Ed. by Paul Guyer and Allen W. Wood. Cambridge and New York: Cambridge University Press.
- Kapitan, Tomis (1999). “The Ubiquity of Self-Awareness.” in: *Grazer Philosophische Studien* 57, pp. 17–43.
- Kaukua, Jari (2015). *Self-Awareness in Islamic Philosophy: Avicenna and Beyond*. Cambridge: Cambridge University Press.
- Keller, Georg B., Tobias Bonhoeffer, and Mark Hübener (2012). “Sensorimotor Mismatch Signals in Primary Visual Cortex of the Behaving Mouse”. In: *Neuron* 74.5, pp. 809–815.
- Kellner, Birgit (2010). “Self-Awareness (Svasamvedana) in Dignāga’s Pramānasamuccaya and -Vṛtti: A Close Reading”. In: *Journal of Indian Philosophy* 38.3, pp. 203–231.
- Killingsworth, Matthew A. and Daniel T. Gilbert (2010). “A Wandering Mind Is an Unhappy Mind”. In: *Science* 330.6006, pp. 932–932.
- King, Lord Peter (1830). *The Life of John Locke*. Vol. II. London: Henry Colburn and Richard Bentley.
- Klatzky, R. (1998). “Allocentric and Egocentric Spatial Representations”. In: *Spatial Information Theory. A Theoretical Basis for GIS, COSIT93* 716, pp. 14–24.
- Knuuttila, Simo and Juha Sihvola (2014). *Sourcebook for the History of the Philosophy of Mind Philosophical Psychology from Plato to Kant*. Dordrecht: Springer.
- Kriegel, Uriah (forthcoming). “The Three Circles of Consciousness”. In: *The Sense of Mineness*. Ed. by M. Guillot and M. Garcia-Carpintero. Oxford: Oxford University Press.
- (2003a). “Consciousness as Intransitive Self-Consciousness: Two Views and an Argument”. In: *Canadian Journal of Philosophy* 33.1, pp. 103–132.
- (2003b). “Consciousness as Sensory Quality and as Implicit Self-Awareness”. In: *Phenomenology and the Cognitive Sciences* 2.1, pp. 1–26.
- (2004). “Consciousness and Self-Consciousness”. In: *The Monist* 87.2, pp. 182–205.

- Kriegel, Uriah (2009). *Subjective Consciousness: A Self-Representational Theory*. Oxford: Oxford University Press.
- (2011). “Précis of Subjective Consciousness: A Self-Representational Theory”. In: *Philosophical Studies* 159.3, pp. 443–445.
- (2013). “Brentano’s Most Striking Thesis: No Representation Without Self-Representation”. In: *Themes from Brentano*. Ed. by Denis Fisette and Guillaume Fréchette. Rodopi. Amsterdam, pp. 22–40.
- (2015). *The Varieties of Consciousness*. Oxford: Oxford University Press.
- Kühle, Lana (2017). “The Missing Pieces in the Scientific Study of Bodily Awareness”. In: *Philosophical Psychology* 30.5, pp. 571–593.
- Lähteenmäki, Vili (2007). “Orders of Consciousness and Forms of Reflexivity in Descartes”. In: *Consciousness. Studies In The History Of Philosophy Of Mind*. Springer, Dordrecht, pp. 177–201.
- Langdon, Robyn and Max Coltheart (2000). “The Cognitive Neuropsychology of Delusions”. In: *Mind & Language* 15.1, pp. 184–218.
- Legrand, Dorothée (2006). “The Bodily Self: The Sensori-Motor Roots of Pre-Reflective Self-Consciousness”. In: *Phenomenology and the Cognitive Sciences* 5.1, pp. 89–118.
- (2007). “Pre-Reflective Self-as-Subject from Experiential and Empirical Perspectives”. In: *Consciousness and Cognition. Subjectivity and the Body* 16.3, pp. 583–599.
- Leibniz, G. W. (1704/1996). *New Essays on Human Understanding*. Trans. by Peter Remnant and Jonathan Bennett. 2nd edition. Cambridge and New York: Cambridge University Press.
- Letheby, Chris and Philip Gerrans (2017). “Self Unbound: Ego Dissolution in Psychedelic Experience”. In: *Neuroscience of Consciousness* 2017.1. eprint: <https://academic.oup.com/nc/article-pdf/2017/1/nix016/25024251/nix016.pdf>.
- Levine, Joseph (2001). *Purple Haze: The Puzzle of Consciousness*. Vol. 6. Oxford University Press Oxford.
- (2007). “Two Kinds of Access”. In: *Behavioral and Brain Sciences* 30.5-6, pp. 514–515.
- Lewis, David (1979). “Attitudes de Dicto and de Se”. In: *Philosophical Review* 88.4, pp. 513–543.
- (1995). “Should a Materialist Believe in Qualia?”. In: *Australasian Journal of Philosophy* 73.1, pp. 140–144.
- Loar, Brian (1990). “Phenomenal States”. In: *Philosophical Perspectives* 4, pp. 81–108.
- Locke, John (1689/1979). *An Essay Concerning Human Understanding*. Ed. by P. H. Nidditch. The Clarendon Edition of the Works of John Locke. Oxford and New York: Clarendon Press.
- Longuenesse, Béatrice (2017). *I, Me, Mine: Back to Kant, and Back Again*. Oxford and New York: Oxford University Press.
- Lycan, William G. (1995). “Consciousness as Internal Monitoring, I: The Third Philosophical Perspectives Lecture”. In: *Philosophical Perspectives* 9, pp. 1–14.
- (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.
- MacKenzie, Matthew D. (2007). “The Illumination of Consciousness: Approaches to Self-Awareness in the Indian and Western Traditions”. In: *Philosophy East and West* 57.1, pp. 40–62.
- (2016). “(Re-)Constructing the Self”. In: *Journal of Consciousness Studies* 23.1-2, pp. 105–124.
- (2017). “Luminous Mind: Self-Luminosity versus Other-Luminosity in Indian Philosophy of Mind”. In: *Indian Epistemology and Metaphysics*. Ed. by Joerg Tuske. London: Bloomsbury Academic, pp. 335–353.
- Maher, Brendan A. (1999). “Anomalous Experience in Everyday Life: Its Significance for Psychopathology”. In: *Monist* 82.4, pp. 547–570.

- Maravita, Angelo (2008). "Spatial Disorders". In: *Cognitive Neurology: A Clinical Textbook*. Ed. by S. F. Cappa et al. New York: Oxford University Press, pp. 89–118.
- Martin, M. G. F. (1995). "Bodily Awareness: A Sense of Ownership". In: *The Body and the Self*. Ed. by Jose Luis Bermudez, Anthony J. Marcel, and Naomi M. Eilan. Cambridge, MA: MIT Press, pp. 267–289.
- McClelland, Tom (forthcoming). "Self-Representationalism". In: *The Oxford Handbook of the Philosophy of Consciousness*. Ed. by Uriah Kriegel. Oxford: Oxford University Press.
- (2015). "Affording Introspection: An Alternative Model of Inner Awareness". In: *Philosophical Studies* 172.9, pp. 2469–2492.
- McGinn, Colin (1991). *The Problem of Consciousness: Essays Toward a Resolution*. Blackwell.
- (1996). *The Character of Mind: An Introduction to the Philosophy of Mind*. Oxford and New York: Oxford University Press.
- Mehta, Neil (2013). "Is There a Phenomenological Argument for Higher-Order Representationalism?" In: *Philosophical Studies* 164.2, pp. 357–370.
- Meier, Barbara (1993). "Speech and Thinking in Dreams". In: *Dreaming as Cognition*. Hertfordshire, UK: Harvester Wheatsheaf, pp. 58–76.
- Metzinger, Thomas (2003). *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.
- Millière, Raphaël (forthcoming). "The Varieties of Selflessness". In: *Philosophy and the Mind Sciences*.
- (2017). "Looking For The Self: Phenomenology, Neurophysiology and Philosophical Significance of Drug-Induced Ego Dissolution". In: *Frontiers in Human Neuroscience* 11.245, pp. 1–22.
- Millière, Raphaël et al. (2018). "Psychedelics, Meditation and Self-Consciousness". In: *Frontiers in Psychology* 9.1475, pp. 1–29.
- Montague, Michelle (2016a). "Cognitive Phenomenology and Conscious Thought". In: *Phenomenology and the Cognitive Sciences* 15.2, pp. 167–181.
- (2016b). *The Given: Experience and Its Content*. Oxford: Oxford University Press.
- Musholt, Kristina (2015). *Thinking about Oneself: From Nonconceptual Content to the Concept of a Self*. Cambridge, MA: The MIT Press.
- Nagel, Thomas (1974). "What Is It Like to Be a Bat?" In: *The Philosophical Review* 83.4, pp. 435–450.
- Nathan, N. M. L. (1997). "Self and Will". In: *International Journal of Philosophical Studies* 5.1, pp. 81–94.
- Neisser, Ulric (1988). "Five Kinds of Self-knowledge". In: *Philosophical Psychology* 1.1, pp. 35–59.
- Nelkin, Norton (1989). "Propositional Attitudes and Consciousness". In: *Philosophy and Phenomenological Research* 49.March, pp. 413–30.
- Nichols, Sarah, Clovissa Haldane, and John R. Wilson (2000). "Measurement of Presence and Its Consequences in Virtual Environments". In: *International Journal of Human-Computer Studies* 52.3, pp. 471–491.
- Nida-Rümelin, Martine (forthcoming). "Experiencing Subjects and So-Called Mine-Ness". In: *The Sense of Mineness*. Ed. by Marie Guillot and Manuel Garcia-Carpintero. Oxford: Oxford University Press.
- (2014). "Basic Intentionality, Primitive Awareness and Awareness of Oneself". In: *Mind, Values, and Metaphysics*. Ed. by Anne Reboul. Springer International Publishing, pp. 261–290.
- (2017). "Self-Awareness". In: *Review of Philosophy and Psychology* 8.1, pp. 55–82.
- (2018). "The Experience Property Frame Work: A Misleading Paradigm". In: *Synthese* 195.8, pp. 3361–3387.

- Noë, Alva (2005). *Action in Perception*. MIT Press.
- O’Keefe, J. and J. Dostrovsky (1971). “The Hippocampus as a Spatial Map. Preliminary Evidence from Unit Activity in the Freely-Moving Rat”. In: *Brain Research* 34.1, pp. 171–175.
- O’Keefe, J. and Lynn Nadel (1978). *The Hippocampus as a Cognitive Map*. Oxford and New York: Oxford University Press.
- O’Shaughnessy, Brian (2000). *Consciousness and the World*. Oxford: Oxford University Press.
- Parfit, Derek (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Parrott, Matthew (2017). “Self-Blindness and Self-Knowledge”. In: *Philosophers’ Imprint* 17.
- Peacocke, Christopher (1992). *A Study of Concepts*. Cambridge, MA: MIT Press.
- (1998). *Being Known*. Oxford: Oxford University Press.
- (2006). “Mental Action and Self-Awareness”. In: *Contemporary Debates in the Philosophy of Mind*. Ed. by Jonathan D. Cohen and Brian P. McLaughlin. Oxford: Blackwell.
- (2014). *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness*. Oxford: Oxford University Press.
- (2016). “The Nature and Role of First and Second Person Content”. In: *Analysis* 76.3, pp. 345–354.
- (2017). “Philosophical Reflections on the First Person, the Body, and Agency”. In: *The Subject’s Matter: Self-Consciousness and the Body*. Ed. by Frédérique de Vignemont and Adrian Alsmith. Cambridge, MA: MIT Press, pp. 289–310.
- (2019). *The Primacy of Metaphysics*. Oxford: Oxford University Press.
- Perry, John (1979). “The Problem of the Essential Indexical”. In: *Noûs* 13.1, pp. 3–21.
- Pitt, David (2004). “The Phenomenology of Cognition Or What Is It Like to Think That P?”. In: *Philosophy and Phenomenological Research* 1, pp. 1–36.
- Price, Matthew et al. (2011). “Does Engagement with Exposure Yield Better Outcomes? Components of Presence as a Predictor of Treatment Response for Virtual Reality Exposure Therapy for Social Phobia”. In: *Journal of Anxiety Disorders* 25.6, pp. 763–770.
- Prinz, Jesse J. (2011). “The Sensory Basis of Cognitive Phenomenology”. In: *Cognitive Phenomenology*. Ed. by Tim Bayne and Michelle Montague. Oxford: Oxford University Press, pp. 174–196.
- (2012). *The Conscious Brain: How Attention Engenders Experience*. Oxford and New York: Oxford University Press.
- Proust, Joelle (2009). “Is There a Sense of Agency for Thought?”. In: *Mental Actions*. Ed. by Lucy O’Brien and Matthew Soteriou. Oxford: Oxford University Press.
- Ramamurthi, B. (1995). “The Fourth State of Consciousness: The Thuriya Avastha”. In: *Psychiatry and Clinical Neurosciences* 49.2, pp. 107–110.
- Reade, Charles (1856). *It Is Never Too Late to Mend: A Matter of Fact Romance*. Vol. 1. Leipzig: Bernard Tauchnitz.
- Recanati, François (2009). “De Re and de Se”. In: *Dialectica* 63.3, pp. 249–269.
- Regenbrecht, Holger and Thomas Schubert (2002). “Real and Illusory Interactions Enhance Presence in Virtual Environments”. In: *Presence: Teleoperators and Virtual Environments* 11.4, pp. 425–434.
- Riva, Giuseppe et al. (2007). “Affective Interactions Using Virtual Reality: The Link between Presence and Emotions”. In: *CyberPsychology & Behavior* 10.1, pp. 45–56.
- Robillard, Geneviève et al. (2003). “Anxiety and Presence during VR Immersion: A Comparative Study of the Reactions of Phobic and Non-Phobic Participants in Therapeutic Virtual Environments Derived from Computer Games”. In: *CyberPsychology & Behavior* 6.5, pp. 467–476.
- Rödl, Sebastian (2007). *Self-Consciousness*. Cambridge, MA: Harvard University Press.

- Romano, Daniele and Angelo Maravita (2019). "The Dynamic Nature of the Sense of Ownership after Brain Injury. Clues from Asomatognosia and Somatoparaphrenia". In: *Neuropsychologia* 132, p. 107119.
- Rosenthal, David M. (1986). "Two Concepts of Consciousness". In: *Philosophical Studies* 49.3, pp. 329–359.
- (1990). *A Theory of Consciousness*. Zentrum für interdisziplinäre Forschung.
- (1993). "Thinking That One Thinks". In: *Consciousness: Psychological and Philosophical Essays*. Ed. by Martin Davies and Glyn W. Humphreys. Blackwell.
- (1997). "A Theory of Consciousness". In: *The Nature of Consciousness*. Ed. by Ned Block, Owen J. Flanagan, and Guven Guzeldere. Cambridge, MA: MIT Press.
- (2005). *Consciousness and Mind*. Oxford: Clarendon Press.
- Rubia, Katya (2009). "The Neurobiology of Meditation and Its Clinical Effectiveness in Psychiatric Disorders". In: *Biological Psychology* 82.1, pp. 1–11.
- Ruby, Florence J. M., Jonathan Smallwood, Haakon Engen, et al. (2013). "How Self-Generated Thought Shapes Mood—The Relation between Mind-Wandering and Mood Depends on the Socio-Temporal Content of Thoughts". In: *PLOS ONE* 8.10, e77554.
- Ruby, Florence J. M., Jonathan Smallwood, Jerome Sackur, et al. (2013). "Is Self-Generated Thought a Means of Social Problem Solving?" In: *Frontiers in Psychology* 4.
- Russell, Bertrand (1926). "Philosophical Consequences of Relativity". In:
- Ryle, Gilbert (1949). *The Concept of Mind*. Hutchinson and Co.
- Sartre, Jean-Paul (1943/1948). *Being and Nothingness*. Trans. by Hazel E. Barnes. New York: Philosophical Library.
- (1936/2004). *The Transcendence of the Ego: A Sketch for a Phenomenological Description*. Trans. by Andrew Brown. London: Routledge.
- Schellenberg, Susanna (2007). "Action and Self-Location in Perception". In: *Mind* 116.463, pp. 603–632.
- (2016). "De Se Content and De Hinc Content". In: *Analysis* 76.3, pp. 334–345.
- Schooler, Jonathan W. et al. (2011). "Meta-Awareness, Perceptual Decoupling and the Wandering Mind". In: *Trends in Cognitive Sciences* 15.7, pp. 319–326.
- Schubert, Thomas, Frank Friedmann, and Holger Regenbrecht (2001). "The Experience of Presence: Factor Analytic Insights". In: *Presence: Teleoperators and Virtual Environments* 10.3, pp. 266–281.
- Schwenkler, John (2014). "Vision, Self-Location, and the Phenomenology of the 'Point of View'". In: *Noûs* 48.1, pp. 137–155.
- Schwind, Valentin et al. (2019). "Using Presence Questionnaires in Virtual Reality". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA: ACM, 360:1–360:12.
- Schwitzgebel, Eric (2007). "Do You Have Constant Tactile Experience of Your Feet in Your Shoes? Or Is Experience Limited to What's in Attention?" In: *Journal of Consciousness Studies* 14.3, pp. 5–35.
- (2016). "Phenomenal Consciousness, Defined and Defended as Innocently as I Can Manage". In: *Journal of Consciousness Studies* 23.11-12, pp. 224–235.
- Scott, William Henry (1918). "Consciousness and Self-Consciousness". In: *The Philosophical Review* 27.1, pp. 1–20.
- Searle, John R. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Sebastián, Miguel Ángel (2012). "Experiential Awareness: Do You Prefer "It" to "Me" ?" In: *Philosophical Topics* 40.2, pp. 155–177.

- Selby, John (2003). *Seven Masters, One Path: Meditation Secrets from the World's Greatest Teachers*. San Francisco: HarperOne.
- Shoemaker, Sydney (1968). "Self-Reference and Self-Awareness". In: *Journal of Philosophy* 65.October, pp. 555–67.
- (1986). "Introspection and the Self". In: *Midwest Studies in Philosophy* 10.1, pp. 101–120.
- (1994a). "Phenomenal Character". In: *Noûs* 28.1, pp. 21–38.
- (1994b). "Self-Knowledge and 'Inner Sense'". In: *Philosophy and Phenomenological Research* 54.2, pp. 249–314.
- Siegel, Susanna (2007). "How Can We Discover the Contents of Experience?" In: *The Southern Journal of Philosophy* 45.S1, pp. 127–142.
- Siewert, Charles (1998). *The Significance of Consciousness*. Princeton, NJ: Princeton University Press.
- (2013). "Phenomenality and Self-Consciousness". In: *Phenomenal Intentionality*. Ed. by Uriah Kriegel. Oxford: Oxford University Press, pp. 235–257.
- Siff, Jason (2014). *Thoughts Are Not the Enemy: An Innovative Approach to Meditation Practice*. Boston: Shambhala Publications Inc.
- Slater, Mel (2009). "Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535, pp. 3549–3557.
- (2018). "Immersion and the Illusion of Presence in Virtual Reality". In: *British Journal of Psychology* 109.3, pp. 431–433.
- Slater, Mel and Maria V. Sanchez-Vives (2016). "Enhancing Our Lives with Immersive Virtual Reality". In: *Frontiers in Robotics and AI* 3.
- Smallwood, Jonathan, Louise Nind, and Rory C. O'Connor (2009). "When Is Your Head at? An Exploration of the Factors Associated with the Temporal Focus of the Wandering Mind". In: *Consciousness and Cognition* 18.1, pp. 118–125.
- Smallwood, Jonathan, Jonathan W. Schooler, et al. (2011). "Self-Reflection and the Temporal Focus of the Wandering Mind". In: *Consciousness and Cognition*. From Dreams to Psychosis: A European Science Foundation Exploratory Workshop 20.4, pp. 1120–1126.
- Smith, David Woodruff (1986). "The Structure of (Self-) Consciousness". In: *Topoi* 5.2, pp. 149–156.
- (2004). *Mind World: Essays in Phenomenology and Ontology*. Cambridge and New York: Cambridge University Press.
- Smith, Joel (2017). "Self-Consciousness". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2017. Metaphysics Research Lab, Stanford University.
- Snow, Michael P. and Robert C. Williges (1998). "Empirical Models Based on Free-Modulus Magnitude Estimation of Perceived Presence in Virtual Environments". In: *Human Factors* 40.3, pp. 386–402.
- Song, Xiaolan and Xiao Wang (2012). "Mind Wandering in Chinese Daily Lives – An Experience Sampling Study". In: *PLOS ONE* 7.9, e44423.
- Sprigge, Timothy L. S. (1969). "The Privacy of Experience". In: *Mind* 78.312, pp. 512–521.
- (1971). "Final Causes". In: *Proceedings of the Aristotelian Society, Supplementary Volumes* 45, pp. 149–192.
- Stawarczyk, David, Helena Cassol, and Arnaud D'Argembeau (2013). "Phenomenology of Future-Oriented Mind-Wandering Episodes". In: *Frontiers in Psychology* 4.
- Stawarczyk, David, Steve Majerus, et al. (2011). "Mind-Wandering: Phenomenology and Function as Assessed with a Novel Experience Sampling Method". In: *Acta Psychologica* 136.3, pp. 370–381.

- Stoljar, Daniel (2016). "The Semantics of 'What It's like' and the Nature of Consciousness". In: *Mind* 125.500, pp. 1161–1198.
- Strawson, Galen (1994). *Mental Reality*. Cambridge, MA: MIT Press.
- (2010). "Radical Self-Awareness". In: *Self, No Self?: Perspectives From Analytical, Phenomenological, and Indian Traditions*. Ed. by Mark Siderits, Evan Thompson, and Dan Zahavi. Oxford University Press.
- (2011a). "Cognitive Phenomenology: Real Life". In: *Cognitive Phenomenology*. Ed. by Tim Bayne and Michelle Montague. Oxford University Press, pp. 285–325.
- (2011b). *Selves: An Essay in Revisionary Metaphysics*. Oxford: Oxford University Press.
- (2013). "Self-Intimation". In: *Phenomenology and the Cognitive Sciences* 14.1, pp. 1–31.
- Sutherland, Norman Stuart, ed. (1989). *The International Dictionary of Psychology*. New York: Continuum.
- Sznycer, Daniel (2019). "Forms and Functions of the Self-Conscious Emotions". In: *Trends in Cognitive Sciences* 23.2, pp. 143–157.
- Thiel, Udo (2011). *The Early Modern Subject: Self-Consciousness and Personal Identity from Descartes to Hume*. Oxford and New York: Oxford University Press.
- Thomasson, Amie L. (2006). "Self-Awareness and Self-Knowledge". In: *Psyche* 12.2.
- Thompson, Evan (2007). *Mind in Life Biology, Phenomenology and the Sciences of Mind*. 1 edition. Cambridge, Mass: Harvard University Press.
- Timmermann, Christopher et al. (2019). "Neural Correlates of the DMT Experience Assessed with Multivariate EEG". In: *Scientific Reports* 9.1, pp. 1–13.
- Titchener, E. B. (1911). "A Note of the Consciousness of Self". In: *The American Journal of Psychology* 22.4, pp. 540–552.
- Tolman, E. C. (1948). "Cognitive Maps in Rats and Men". In: *Psychological Review* 55.4, pp. 189–208.
- Town, Stephen M., W. Owen Brimijoin, and Jennifer K. Bizley (2017). "Egocentric and Allocentric Representations in Auditory Cortex". In: *PLOS Biology* 15.6, e2001878.
- Troje, Nikolaus F. (2019). "Reality Check". In: *Perception* 48.11, pp. 1033–1038.
- Tsakiris, Manos (2011). "The Sense of Body Ownership". In: *The Oxford Handbook of the Self*. Ed. by Shaun Gallagher. Oxford: Oxford University Press.
- Tye, Michael and Briggs Wright (2011). "Is There a Phenomenology of Thought?" In: *Cognitive Phenomenology*. Ed. by Tim Bayne and Michelle Montague. Oxford: Oxford University Press, pp. 326–344.
- Usoh, Martin et al. (2000). "Using Presence Questionnaires in Reality". In: *Presence: Teleoperators and Virtual Environments* 9.5, pp. 497–503.
- Vallar, Giuseppe and Roberta Ronchi (2009). "Somatoparaphrenia: A Body Delusion. A Review of the Neuropsychological Literature". In: *Experimental Brain Research* 192.3, pp. 533–551.
- Van Baren, Joy and Wijnand IJsselstein (2004). *Measuring Presence: A Guide to Current Measurement Approaches*. OmniPres.
- Ward, Jamie and Michael J. Banissy (2015). "Explaining Mirror-Touch Synesthesia". In: *Cognitive Neuroscience* 6.2-3, pp. 118–133.
- Weech, Séamas, Sophie Kenny, and Michael Barnett-Cowan (2019). "Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review". In: *Frontiers in Psychology* 10.
- Weinberg, Shelley (2016). *Consciousness in Locke*. Oxford: Oxford University Press.
- Weisberg, Josh (2011). "Misrepresenting Consciousness". In: *Philosophical Studies* 154.3, pp. 409–433.
- White, Stephen L. (1987). "What Is It Like to Be a Homunculus?" In: *Pacific Philosophical Quarterly* 68. June, pp. 148–74.

- Wider, Kathleen (2006). "Emotion and Self-Consciousness". In: *Self-Representational Approaches to Consciousness*. Ed. by Uriah Kriegel and Kenneth Williford. Cambridge, MA: MIT Press, pp. 63–87.
- Williamson, Timothy (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Windt, Jennifer M. (2015). *Dreaming: A Conceptual Framework for Philosophy of Mind and Empirical Research*. Cambridge, MA: MIT Press.
- Wirth, Werner et al. (2007). "A Process Model of the Formation of Spatial Presence Experiences". In: *Media Psychology* 9.3, pp. 493–525.
- Witmer, Bob G. and Michael J. Singer (1998). "Measuring Presence in Virtual Environments: A Presence Questionnaire". In: *Presence: Teleoperators and Virtual Environments* 7.3, pp. 225–240.
- Wittgenstein, Ludwig (1958). *The Blue and Brown Books*. Oxford: Blackwell.
- (1947/1980). *Remarks on the Philosophy of Psychology*. Ed. by G. E. M. Anscombe and G. H. von Wright. Trans. by G. E. M. Anscombe. Vol. 1. Oxford: Blackwell.
- Wrigley, William J. and Stephen B. Emmerson (2013). "The Experience of the Flow State in Live Music Performance". In: *Psychology of Music* 41.3, pp. 292–305.
- Wu, Tina L. Y. et al. (2019). "The Effect of Head Tracking on the Degree of Presence in Virtual Reality". In: *International Journal of Human-Computer Interaction* 35.17, pp. 1569–1577.
- Wu, Wayne (forthcoming). "Mineness and Introspective Data". In: *The Sense of Mineness*. Ed. by M. Garcia-Carpintero and M. Guillot. Oxford: Oxford University Press.
- Zahavi, Dan (forthcoming). "Consciousness and Selfhood: Getting Clearer on For-Me-Ness and Mineness". In: *The Oxford Handbook of the Philosophy of Consciousness*. Ed. by Uriah Kriegel. Oxford: Oxford University Press.
- (1999). *Self-Awareness and Alterity: A Phenomenological Investigation*. Evanston, IL: Northwestern University Press.
- (2005). *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, MA: MIT Press.
- (2011). "The Experiential Self: Objections and Clarifications". In: *Self, No Self? Perspectives from Analytical, Phenomenological, and Indian Traditions*. Ed. by Mark Siderits, Evan Thompson, and Dan Zahavi. Oxford: Oxford University Press, pp. 56–78.
- (2014). *Self and Other: Exploring Subjectivity, Empathy, and Shame*. Oxford: Oxford University Press.
- (2017). "Thin, Thinner, Thinnest: Defining the Minimal Self". In: *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*. Ed. by Christoph Durt, Thomas Fuchs, and Christian Tewes. Cambridge, MA: MIT Press, pp. 193–200.
- (2018). "Consciousness, Self-Consciousness, Selfhood: A Reply to Some Critics". In: *Review of Philosophy and Psychology* 9.3, pp. 703–718.
- Zahavi, Dan and Uriah Kriegel (2016). "For-Me-Ness: What It Is and What It Is Not". In: *Philosophy of Mind and Phenomenology: Conceptual and Empirical Approaches*. Ed. by Daniel O. Dahlstrom, Andreas Elpidorou, and Walter Hopp. London: Routledge.
- Zimmerman, Aaron Zachary (2006). "Basic Self-Knowledge: Answering Peacocke's Criticisms of Constitutivism". In: *Philosophical Studies* 128.2, pp. 337–379.
- Zmarz, Pawel and Georg B. Keller (2016). "Mismatch Receptive Fields in Mouse Visual Cortex". In: *Neuron* 92.4, pp. 766–772.