

ACTA PHILOSOPHICA GOTHOBURGENSIA

19

INTROSPECTING  
REPRESENTATIONS

Susanna Radovic



ACTA UNIVERSITATIS GOTHOBURGENSIS

INTROSPECTING  
REPRESENTATIONS

ACTA PHILOSOPHICA GOTHOBURGENSIA

19

# INTROSPECTING REPRESENTATIONS

Susanna Radovic



ACTA UNIVERSITATIS GOTHOBURGENSIS

© Susanna Radovic 2005

Distribution:  
ACTA UNIVERSITATIS GOTHOBURGENSIS  
Box 222  
SE-405 30 Göteborg  
Sweden

ISBN 91-7346-540-2  
ISSN 0283-2380

Printed in Sweden by  
Kompendiet, Göteborg 2005

ACTA PHILOSOPHICA GOTHOBURGENSIA

ISSN 0283-2380

Editors: Ingemar Persson and Dag Westerståhl

Published by the Department of Philosophy of the University of Göteborg

Subscription to the series and orders for single volumes should be addressed to:

ACTA UNIVERSITATIS GOTHOBURGENSIS

Box 222, SE-405 30 Göteborg, Sweden

VOLUMES PUBLISHED

1. MATS FURBERG, THOMAS WETTERSTRÖM and CLAES ÅBERG (editors): Logic and Abstraction. Essays dedicated to Per Lindström on his fiftieth birthday. 1986. 347 pp.
2. STAFFAN CARLSHAMRE: Language and Time. An Attempt to Arrest the Thought of Jacques Derrida. 1986. 253 pp.
3. CLAES ÅBERG (editor): Cum Grano Salis. Essays dedicated to Dick A. R. Haglund. 1989. 263 pp.
4. ANDERS TOLLAND: Epistemological Relativism and Relativistic Epistemology. Richard Rorty and the possibility of a Philosophical Theory of Knowledge. 1991. 156 pp.
5. CLAES STRANNEGÅRD: Arithmetical realizations of modal formulas. 1997. 100 pp.
6. BENGT BRÛLDE: The Human Good. 1998. 490 pp.
7. EVA MARK: Självbilder och jagkonstitution. 1998. 236 pp.
8. MAY THORSETH: Legitimate and Illegitimate Paternalism in Polyethnic Conflicts. 1999. 214 pp.
9. CHRISTIAN MUNTHE: Pure Selection. The Ethics of Preimplantation Genetic Diagnosis and Choosing Children without Abortion. 1999. 310 pp.
10. JOHAN MÅRTENSSON: Subjunctive Conditionals and Time. A Defense of a Weak Classical Approach. 1999. 212 pp.
11. CLAUDIO M. TAMBURRINI: The 'Hand of God'? Essays in the Philosophy of Sports. 2000. 167 pp.
12. LARS SANDMAN: A Good Death: On the Value of Death and Dying. 2001. 346 pp.
13. KENT GUSTAVSSON: Emergent Consciousness. Themes in C.D. Broad's Philosophy of Mind. 2002. 204 pp.
14. FRANK LORENTZON: Fri vilja? 2002. 175 pp.
15. JAN LIF: Can a Consequentialist Be a Real Friend? (Who Cares?). 2003. 167 pp.
16. FREDRIK SUNDQVIST: Perceptual Dynamics. Theoretical Foundations and Philosophical Implications of Gestalt Psychology. 2003. 261 pp.
17. JONAS GREN: Applying Utilitarianism. The Problem of Practical Action-guidance. 2004. 160 pp.
18. NIKLAS JUTH. Genetic Information – Values and Rights. The Morality of Presymptomatic Genetic Testing. 2005. 459 pp.
19. SUSANNA RADOVIC: Introspecting Representations. 2005. 200 pp.

ISBN 91-7346-540-2



## Acknowledgements

Several persons have contributed to the completion of this thesis in different ways. First and foremost I am indebted to my supervisor Helge Malmgren whose insightful suggestions and comments on all levels have been a tremendous help.

I would also like to express special thanks to Filip Radovic for detailed comments on an earlier draft of the manuscript and for interesting philosophical discussions and support throughout the years. Jan Almäng, Björn Haglund and Dag Westerståhl have also contributed with helpful comments and suggestions on an earlier draft of the thesis. I would also like to thank my friends and colleagues at the Philosophy Department in Göteborg. In particular Petra Andersson and Jan Lif and the participants in the dagsem-seminar who in different ways have helped me complete this thesis.

Mattias Dahl has provided me with lay-out and typographical assistance and Kerstin Dahlén has corrected my English. Special thanks to Gulli Andersson who has kept me going by serving all those first-rate meals and finally to Ulf Börjesson for love and support and for taking such good care of my animals when I was elsewhere occupied.

# Contents

<b>Acknowledgements</b>	i
<b>Contents</b>	ii
<b>INTRODUCTION</b>	1
The problem	2
The history	4
Overview of this thesis	7
A preliminary characterisation of introspection	9
<b>PART ONE:</b>	
<b>A BRIEF ENCOUNTER WITH THE PHILOSOPHY OF INTROSPECTION</b>	13
<b>1. The object of introspection</b>	15
1.1 The object of introspection is what is in the mind	15
1.2 The object of introspection is processes in the brain	15
1.3 The object of introspection is the sensory aspect of mental states	17
1.4 The object of introspection is the intentional content of mental states	20
1.5 The object of introspection is behavioural dispositions	21
1.6 Summary of chapter one	23
<b>2. The epistemic status of introspection</b>	24
2.1 Is introspection a reliable method?	24
2.2 First person authority	28
2.2.1 Infallibility	28
2.2.2 Indubitability	33
2.2.3 Incorrigibility by another party	35
2.2.4 Transparency	37
2.2.5 Immediacy	39
2.2.6 Are first person beliefs justified?	43
2.3 First person authority and externalism	46
2.3.1 Phenomenal externalism	50
2.4 “First person reports” are not introspective reports	53
2.5 Summary of chapter two	54

<b>3. The introspective process</b>	56
3.1 Perceptual awareness	56
3.2 Four kinds of inner perception	57
3.2.1 Awareness of awareness	58
3.2.2 Direct and indirect inner perception	62
3.3 Inner thinking	63
3.4 Displaced perception	64
3.5 Replay of perception	64
3.6 Summary	66
3.7 Final comments	66

**PART TWO:**  
**REPRESENTATION, META-REPRESENTATION AND INTROSPECTION** 69

<b>4. Representational theories of consciousness and introspection</b>	72
4.1 Phenomenal consciousness	74
4.2 Qualia as representational properties	77
4.2.1 What is being represented?	80
4.2.2 How do our experiences represent properties of objects?	87
4.3 Higher order theories	89
4.3.1 Higher order perception	90
4.3.2 Higher order thinking	92
4.3.3 Higher order dispositional thinking	94
4.4 One level theories of consciousness	95
4.4.1 Dretske's theory of introspection	96
4.4.2 Tye	99
4.4.3 Shoemaker	100
4.5 Summary of chapter four	102
<b>5. How many levels do we need?</b>	103
5.1 Worldly and experiential subjectivity	103
5.2 Conscious states	109
5.2.1 Unconscious conscious states	112
5.3 Summary of chapter five	115
<b>6. Against HOP</b>	117
6.1 The alleged introspective organ	117
6.2 Special phenomenology	121
6.3 Immunity to error	123
6.3.1 Misidentifying the content and attitude of a mental state	124
6.3.2 Misidentifying the self	125

6.3.2.1 Experienced self-identification	128
6.3.2.2 The self and the world	129
6.3.3 Misidentifying the phenomenal character of an experience	131
6.4 Immunity to ignorance	132
6.5 Distortion of the object	135
6.6 Higher order mis-perceptions	137
6.7 Watching representations	140
6.8 Summary of chapter six	145
<b>7. Against HOT</b>	147
7.1 Phenomenology	147
7.2 Attention	148
7.3 The content of thoughts and perceptions	149
7.4 The relation between a HOT and a LOS	153
7.5 The cons of HOT(d)	156
7.6 Summary of chapter seven	158
<b>8. Against OL</b>	159
8.1 Introspecting hallucinations	159
8.2 Introspecting pains	162
8.3 Justification of introspective beliefs	164
8.3.1 Dretske's view(s) on justification of introspective beliefs	165
8.3.2 Tye, Shoemaker and the recent Dretske	170
8.4 Summary of chapter eight	174
<b>9. Conclusions</b>	176
9.1 HOT	176
9.2 HOT(d)	177
9.3 HOP	178
9.4 Dretske	182
9.5 Tye	185
9.6 Shoemaker	186
9.7 Final remarks	187
<b>Summary (in Swedish)</b>	189
<b>References</b>	192

# Introduction

## THE PROBLEM

During the last couple of decades, so called representationalist theories of mind have gained increased popularity. These theories describe mental states in terms of representations of external objects and states of affairs. It is also often held that the content of a subject's thoughts and perceptions is determined by facts outside her mind, such as social relations between her and other people and causal relations between her and external objects. Some representationalists even argue that the *phenomenal character* of perceptual experiences is determined by external factors in the sense that the truth conditions of statements like: "it looks blue" involve such facts. This entails that so called "phenomenal properties" such as colours are not properties of my experiences or even determined by such properties. This thesis has been labelled "phenomenal externalism" by e.g., Fred Dretske<sup>1</sup> and William Lycan<sup>2</sup>.

Introspection has traditionally been described as a subject's immediate awareness of her own experiences. It has been assumed that the subject has a special and privileged access to her experiences which means that she cannot be mistaken either about the content of her beliefs and experiences or about what they feel like to her. A long lived theory about introspection is that the introspective process is similar to perception, only the objects of the introspective process are "inner" instead of "outer". This model seems to entail that experiences also share relevant similarities with external objects, such as having intrinsic properties, properties the subject is *aware of* when observing the objects in question.

This view of introspection is difficult to combine with phenomenal externalism. For one thing it seems to follow from phenomenal externalism that the subject does not have epistemically secure access to the content of her own experiences since the content is determined by external factors. If she does not have infallible knowledge about the external world, how could she have it about her perceptions and thoughts about the world? Secondly, it seems difficult to analyse introspection in terms of inner perception when experiences cannot be conceived of as objects with observable properties. If an experience is a representation whose content is determined by how things

---

<sup>1</sup> Dretske 1996.

<sup>2</sup> Lycan 2001c.

stand in the world, it does not seem fruitful to look at the representation in order to learn what it is about.

A few contemporary philosophers have set out to solve these problems. My contribution here does not entail a solution to them, only an attempted analysis of how and why these other efforts have failed to present such a solution. Some of these authors still analyse introspection in terms of some kind of inner perception, despite the fact that they think that phenomenal character is determined by represented properties of objects and not by some intrinsic qualities of the experiences themselves. William Lycan is one of these philosophers. Lycan also claims that the existence of inner perception can explain the fact that some mental states are conscious while others are not. A conscious experience is an experience that the subject is *aware of* by inner perception. David Rosenthal attempts a similar solution. He claims that the subject becomes conscious of her own mental states by *thinking* or *believing* that she is in them. Peter Carruthers claims that the subject does not need to actually entertain a thought about a certain experience of hers in order to be conscious of it. It is enough that she is disposed to do so. Mental states that are available to higher order thoughts are phenomenally conscious, i.e., the subject has introspective access to what these states are like to be in.

Other representationalist philosophers are highly critical to this model. They challenge both the assumption that phenomenal consciousness should be accounted for in terms of the subject being introspectively aware of her mental states, and the thesis that introspection is similar to perception in any interesting sense. These authors try to give other accounts of introspective awareness that take the representationalist thesis seriously. If the properties we are aware of when being in a certain mental state are properties of the represented objects, it follows that we are not directly aware of the mental states through introspection, they hold. Fred Dretske, Michael Tye and Sydney Shoemaker attempt versions of such a theory of introspection.

In my view, neither of the theories mentioned here are able to give a plausible account of introspection given these prerequisites. In this thesis, I will try to elaborate this claim.

## THE HISTORY

“The word introspection need hardly be defined – it means, of course, the looking into our own minds and reporting what we there discover.”<sup>3</sup>

Maybe this was true in the days of William James and his contemporaries, but it is far from true today. These days no one seems to know what introspection is, or rather they may know each individually, but there is no general agreement amongst philosophers. Back in the days of William James there was no question about whether introspection was anything but a kind of perception-like direct awareness of our own experiences. So what happened?

There are two periods in the history of philosophy of psychology when introspection has been thoroughly discussed: at the end of 19<sup>th</sup> century and the beginning of the 20<sup>th</sup>, and at the end of the 20<sup>th</sup> century and the beginning of the 21<sup>st</sup>. (So maybe it is a turn of the century phenomenon to be preoccupied with this question). There are however different problems that have been in the focus during these two periods. In the beginning of the last century the worries primarily concerned whether introspection was a reliable scientific method or not. The main emphasis in the contemporary discussion lies not on whether introspection is useful as a scientific method, but on its nature.

In 1986 William Lyons wrote: “In philosophy, and in psychology since behaviourism, there have rarely been extended studies of introspection”.<sup>4</sup> Neither does this stand true today. During the last 10 years a number of books has been published that are all about the problem of introspection. *Knowing our Own Minds* (Wright, Smith and Macdonald) from 1998, *New Essays on Semantic Externalism and Self-Knowledge* (Nuccetelli) from 2003, *Privileged Access: Philosophical Theories of Self-Knowledge* (Gertler), also from 2003, and *Introspection Vindicated* by Gregg Ten Elshof from 2005, just to name some. Apart from these books, a vast amount of articles has been published in this area. Some of these publications deal with the same problems that will be discussed in this thesis, as e.g., whether an account of introspection entails two or more levels of awareness, and whether introspection is best described as a kind of inner perception or not. Others articles and books are about related problems which have to do with introspection and self-

---

<sup>3</sup> James 1950 [1890] I, p. 185.

<sup>4</sup> Lyons 1986, p. xi

knowledge, such as how to account for self-reference or the specific problem of combining the thesis that first person knowledge is authoritative with the assumption that mental states are individuated by external circumstances.

Why is there such a growing interest in this subject area today? If one looks upon the history of philosophy of mind one notices that as long as Cartesian dualism was the state of the art theory, philosophers did not concern themselves much with the problem of introspection probably because they did not recognise such a problem in the first place. According to Cartesian dualism there is no conflict between the metaphysical definition of the mind and the epistemological characterisation of self-knowledge. The mind is defined as an immaterial substance, with the capacity of thinking, and something of which the subject has immediate and infallible knowledge. Mental states are taken to have certain properties that are accessible by introspection only, properties that later were to be named “qualia”.

Cartesian dualism as well as the view of qualia as intrinsic properties of our experiences which determine what it is like to have these experiences grew out of fashion. Some new metaphysical claims were being made in their place. The mind was now taken to be determined by its *functional* properties, or by its *neurophysiological* properties and from those perspectives, experiences could not longer be conceived of as having *immaterial* features. What it feels like to have a certain experience was instead taken to be determined by the functional role or the intentional content of the experience. Given these new ideas about the mind, introspection becomes harder to account for. If the mind is not, in principle, different from the external world, what reason do we have to believe that a person’s knowledge about her own mind is special in some way? It may still be intuitively appealing to claim that a person has privileged access to the contents of her own mind, but the explanation hereof must be different from the Cartesian one. It cannot be due to the fact that mental properties are of a certain metaphysical kind to which introspection is attuned.

With the founding of experimental psychology, introspection temporarily became a subject matter of its own. The philosophers before had taken for granted that introspection is similar to perception, only we “look” inwards instead of outwards. The thought that the introspective process in important aspects resembles perception had been around for quite some time, at least since John Locke. The psychologists at the previous turn of the century started to worry about whether consciousness really could be split into two parts, which this perceptual thesis seems to imply. It appears as if

one part of consciousness must be observing another part. Psychologists and philosophers were also concerned about the risk that the inner perceptual process could distort the experience it observes. If that were the case, introspection would after all not be a reliable method for the scientific study of the mind. The behavioural psychologists that followed concluded that introspection is nothing but silent speech. Gilbert Ryle argued that what has been referred to as “introspection” is just to reflect upon one’s own past actions. We do not in any sense of the word, engage ourselves in inner perception.

Introspection is an interesting topic in its own right. How do we know our own minds? Do we have infallible knowledge about our own beliefs and feelings or are we just more certain about these matters than about other things? Is introspection an inner counterpart to perception? If so, to perception in what sense modality and in which sense? Is it similar to vision or does it have more in common with touch? Is it what philosophers have called “direct perception” or is it “indirect perception”? Is introspection some kind of attention? Reflection? Is the result of an introspection always a belief, or is it sometimes just a kind of non-epistemic awareness? The aim of this thesis is not to give definite answers to these questions, but they will all be touched upon since a thorough discussion of introspection must deal with all these issues.

Introspection should also be of great interest for those who are interested in any area of the philosophy of mind. Almost every philosophical book about consciousness that has been published during the last fifteen years deals with introspection in some aspect. It is difficult to completely disregard self-awareness while writing about consciousness, regardless of what the main theme of the text is. Furthermore, since introspection is often being compared to perception, attention, thinking and so forth, theories about introspection are relevant for discussions of these topics as well.

The question of whether introspection is a reliable process or not was, as previously mentioned, widely discussed in the beginning of the 20<sup>th</sup> century. It is interesting to note that after the dethronement of behavioural psychology, introspection is once again used in psychological research. Studies of e.g., meta-cognition seem to involve introspection since meta-cognition entails reflecting on one’s own cognitive capacities. Since the cognitive model of the mind basically is representationalist as well, the subject of this thesis should also be of some interest to cognitive psychologists.

## OVERVIEW OF THIS THESIS

The thesis consists of two parts. In the first part I will attempt to give a general overview of theories of introspection by describing how philosophers have accounted for, in turn; the object of introspection, the epistemic status of introspective beliefs, and the introspective process as such. The reader will see examples of both the “traditional” view of introspection where the term stands for the subject’s privileged access to the intrinsic phenomenal character of her own mental states, as well as of the analysis of introspection in terms of indirect awareness of experiences that lack intrinsic properties, which is the view point of most of the theories discussed in the second part of the thesis.

The first chapter will concern views about the nature of the putative objects of introspection. Some philosophers have argued that the object of introspection is the sensory aspect of experiencing. The idea is that the subject has introspective access only to what it is like to have certain experiences or entertain certain thoughts. For example, assume that I am looking at a blue coffee mug. By introspecting this experience I may become aware of how the mug *appears to me*, or *what it is like* for me to look at the mug. According to this view I have on the other hand no introspective access to the content of my visual experience, that it is *a coffee mug* I see. Others, again, have argued that the subject only has introspective access to the intentional content of her experiences. But they also often hold that what it is like for me to look at e.g., a blue mug is entirely a matter of what my experience represents.

The second chapter will concern the epistemic considerations mentioned above. Most philosophers think that self-knowledge is authoritative in some sense. But while some argue that introspective knowledge is infallible, others just claim that the subject is in a slightly better position than other people when it comes to judging what is going on in her mind. The question of first person authority is connected to that of whether introspection is direct or not. And again, different philosophers put different meanings to the term “direct”. Finally, a challenge which has been put forward to authoritative self-knowledge, namely semantic externalism will be briefly discussed. If knowing what we think about entails knowing what the world is like, does this mean that we have a priori knowledge about the external world, or does it mean that we can not be certain about what we think about while thinking it?

The third chapter of this introductory discussion of introspection concerns the nature of the introspective process. A popular view in the history of philosophy of mind is that introspection in important aspects is similar to perception. Since theories about the nature of perception are diverging, so are the views about its inner counterpart. The introspective process has also been described as being more similar to thinking than to perceiving.

In the second part of the thesis I will first briefly describe what the representationalist theory of mind entails, and especially when it is combined with phenomenal externalism. I will also try to describe how the different theories are different from one another. I will distinguish between on the one hand: higher order theories of consciousness (HO-theories) and, on the other, one-level theories (OL-theories). The first kind of theory entails that introspection is a higher order awareness that is either perceptual or like thinking. One-level theories, on the other hand maintain that phenomenal consciousness need not be explained in terms of different orders of awareness. It should instead be explained by the function conscious states play in the cognitive system. Since these theories go against the description of introspection as a kind of direct awareness of mental states, they will instead limit the scope of introspection to only entail the subject's *awareness that* she has an experience with a certain intentional content.

This discussion has attracted many interested authors. I will however limit the discussion to the main characters, those who have initiated the debate and those to which all other writers in the debate refer. I.e., on the one side: David Armstrong, Peter Carruthers, William Lycan, and David Rosenthal and on the other: Fred Dretske, Sydney Shoemaker and Michael Tye.

Some of the theories that will be reviewed here do not presuppose phenomenal externalism. Rosenthal claims that the phenomenal character of experiencing is given by intrinsic properties of the experience itself. Carruthers argues that this character should be understood as part of the representational content, but that this type of content is narrow instead of wide. Shoemaker argues in favour of a one-level theory that assumes that perceptual states represent both objective and phenomenal properties of external objects which means that he holds a representationalist theory that is, in a sense, both externalist and internalist. Armstrong, finally, claims that mental states are determined by their causal roles.

In the chapters that follows, (6, 7 and 8) arguments that have been held against the different theories will be considered. I will also present some

of my own. The conclusion is that none of these attempts to give a positive account of introspection are successful. The HO-theorists generally fail to give a comprehensible account of how we can be directly aware of the represented properties of mental states. The OL-theories, on the other hand, fail to give any positive account of introspective awareness. Self-awareness becomes limited to awareness that one has an experience with a certain conceptual content. This means that they can, e.g., not account for introspective awareness of perceptual states that in themselves do not, in themselves, entail conceptual awareness. If I have a perceptual experience of a power takeoff while not knowing that it is a power takeoff I see, it follows from these theories that I will not be able to form an introspective belief about this perceptual experience. Neither can they produce a story that explains how a subject from an introspective point of view is able to distinguish between different modes of experiencing, such as whether an experience of mine is visual or auditory.

It needs to be added that representationalism regarding mental states has been contested by several other philosophers who argue that the qualitative character of experiences is not determined by the representational content. One of the most hard-working critics of representationalism is Ned Block.<sup>5</sup> The aim of this thesis is not primarily to address the question whether representationalism is accurate or inaccurate. In my view, representationalism combined with phenomenal externalism is an appealing theory. The fact that none of the theories discussed here can account for introspection given this metaphysical background does not prove that no such theory is true. It is indeed a worthy task to continue to find a way to combine these theses.

## A PRELIMINARY CHARACTERISATION OF INTROSPECTION

Before we embark on an exposé of different views on what introspection is, it might be fruitful to give a preliminary characterisation of introspection, so we at least have a slight idea of what we are talking about. This account will have to meet two constraints: it should capture the most common ideas among philosophers about the nature of introspection, and it should not

---

<sup>5</sup> Block, e.g., 1998, 2003.

predetermine too many answers to interesting philosophical questions, some of which we are going to deal with later.

Few attempts have been made in the last decades to give clear-cut definitions of the term “introspection”, yet almost every philosopher of mind discusses the phenomenon of introspection. As I mentioned earlier, opinions about the nature of introspection and introspective knowledge are diverging, yet there seems to be some basic and widely accepted ideas about what introspection is, even if they are not always made explicit by the writers themselves. Most philosophers agree that the result of an introspective attempt is some kind of awareness about something which goes on in the subject’s *own mind*. It is also fairly accepted that the introspective process, if there is such, is *inner* in some important sense. Each person must have arrived at her introspective conclusions by some other means than by ordinary external perception. If I come to believe that I like horses more than pigs just because a trusted friend tells me so, this belief will not count as an introspective belief. Many philosophers who write about introspection also seem to think that the scope of introspection is limited to *occurrent* mental states. When we introspect something, that something is already present to our minds. There are some exceptions though, as e.g., Ryle.<sup>6</sup> It is perhaps a matter of taste whether the characterisation of introspection given here should be widened to also capture Ryle’s account of introspection, or if Ryle’s conception of introspection should be regarded as atypical. I am leaning towards the second suggestion. We will return to Ryle’s theory in subsection 1.5.

It is important to keep in mind that not everything that goes on “in our heads” should count as introspecting, just because it is “inner”. Head calculations, deliberations on what actions one should take in a certain situation, attempts to remember the name of an acquaintance, and daydreaming are all examples of “inner” episodes which can easily be confused with introspection. It is not introspecting when I try to remember the name of an acquaintance, that is remembering. And I do not introspect when I try to envision a possible outcome of tomorrow’s riding tour. It might be argued that to introspect is nothing but to perform some of these other mental actions, but we could at least try to give a sense to

---

<sup>6</sup> According to Ryle (1949) knowledge of our own personality traits is a kind of introspective knowledge, maybe even the only kind. But probably, even he would argue that my introspective knowledge of my personality must be based on my own observations of my own behaviour, and not on suggestions made to me by other people about my behaviour.

“introspection” which does not coincide entirely with the meaning of another mental term.

One way to make sure that the characterisation of introspection is informative and that the description of introspection does not coincide with that of another mental process is to demand that the subject through this alleged process becomes aware of something “new”. Assume that a person perceives a green apple. She is then aware of a green apple and if she is also aware that the apple is green she will be able to judge: “the apple is green”. Let us then also assume that the subject introspects her apple perception. She should then become aware of something new and different, perhaps of the fact that she perceives a green apple. In sum:

(i) Introspective awareness is awareness of the subject’s *occurrent mental states*. The state in question does however not need to be episodic. A perception e.g., can last for a long period of time. Someone may feel an intense pain, and nothing but the pain for several hours. There are also other possible objects of introspection that go on for a considerable time, like e.g., a feeling of depression.

(ii) The introspective process is inner in some sense. It is not identical with any of the perceptual processes that give us information about the external world including our own body.

(iii) Introspection should make the subject aware of something she was not previously aware of. Specifically, the belief that the apple is green and the belief that I see a green apple should be distinct states in so far as they should have distinct truth conditions.



Part one:

a brief encounter  
with the philosophy  
of introspection

In the recent history of philosophy there has been considerable disagreement regarding both the *scope* of introspection and its *epistemic* status, as well as about the nature of the *introspective process*. The majority of philosophers argue that only occurrent mental states, in contrast to dispositional ones, can be introspected, and a few philosophers claim that the introspective scope is limited to the sensory aspect of a person's mental acts. There are also divergent opinions regarding the epistemic status of introspection. Some philosophers claim that it is *impossible* to be wrong about the content as well as about the sensory or qualitative aspect of one's own mental states. A slightly weaker position claims that a person's reports about her own states bear a special authority, even though they are not infallible. Some philosophers have argued that introspective knowledge is only as certain as any other kind of empirical knowledge, and some philosophers have maintained that introspective knowledge is less reliable than knowledge of extra-mental happenings. Finally, the position has also been upheld that the deliverances of introspection do not comprise knowledge at all.

The most elusive aspect of introspection seems to be the nature of the process as such. What is it that we do (if anything) when we introspect? Is introspection necessarily an activity or can we be introspectively aware of something without actually doing anything? Can the introspective process successfully be compared to perception, and if so, perception in what modality and in what sense? In case the answer is "vision", which kind of visual perceptual process provides the best metaphor? Do we *look* at mental events, do we *see* them, do we seem them *as* something, do we see *that* they are so and so?

The high-level survey of theories of introspection which follows is intended as a background to the main issue of this thesis; *whether introspection can intelligibly be construed as awareness of mental representations*. The theories discussed in the rest of this part are chosen on the basis of having relevance for a proper understanding of the contemporary theories discussed in the second part.

# 1. The object of introspection

First, different views on *what it is* we introspect will be considered.

## 1.1 THE OBJECT OF INTROSPECTION IS WHAT IS IN THE MIND

The object of introspection is easily delimited if you ascribe to a Cartesian world view. The world consists of two different substances: mind and matter. The object of introspection is the mind or rather what is *in* the mind. To introspect means, etymologically, *to look inwards*, and for a Cartesian dualist “inner” means (sometimes) in the mind. So in this framework, the object of perception is the external world, while that of introspection is its internal counterpart - the mind.

The mind is defined as immaterial and something that thinks.<sup>7</sup> Touching one of your teeth with the tip of your tongue is, hence, not introspection, since the tooth is a material object and therefore does not belong to the mind. A thinking computer that scans its own operations would not be attributed introspective abilities either, since the computer and its “thinking” operations are both objects of matter. If we define the object of introspection in this way, we are at no great risk of confusing introspection with for instance perception; it gets more complicated if we want to uphold a non-dualist metaphysics.

## 1.2 THE OBJECT OF INTROSPECTION IS PROCESSES IN THE BRAIN

The term “introspection” is also used in a similar way by philosophers who are not in favour of a Cartesian world view. “To introspect” still means “to look inwards”, according to some materialist philosophers, but “inner” is

---

<sup>7</sup> Descartes mentions doubting, understanding, affirming, denying, willing, refusing, imaging and feeling as examples of thinking. (Meditation II. 1973 [1911].)

used in a different sense here. The mind is “inner” only in the sense that it is located *inside the body*. Paul Churchland writes:

”[O]ne’s introspective consciousness of oneself appears very similar to one’s perceptual consciousness of the external world. The difference is that, in the former case, whatever mechanisms of discrimination are at work are keyed to internal circumstances instead of external ones. [...] Self-consciousness is no more (and no less) mysterious than perception generally. It is just directed internally rather than externally.”<sup>8</sup>

The difference between perceiving and introspecting seems according to Churchland to be a question of *where* we are “looking” (and, of course, what we are looking with). Through the introspective process we examine what is inside the boundaries of our bodies. A possible objection is that the fact that a given process takes place inside the body is not sufficient for its being qualified as a mental process. To touch one of your teeth with your tongue does not count as introspection and your tooth is after all inside your body. One common move made by materialists in the face of this problem is to delimit introspection from perception by claiming that in introspection we attend to the happenings in our own brain (or the cortex, or certain parts of the cortex). Churchland again:

“Dopamine levels in the limbic system, the spiking frequencies in specific neural pathways, resonances in the *n*th layer of occipital cortex, inhibitory feedback to the lateral geniculate nucleus, and countless other neurophysiological niceties could be moved into the objective focus of our introspective discrimination, just as *Gm7* chords and *Adim* chords are moved into objective focus of a trained musician’s auditory discrimination.”<sup>9</sup>

Michael Levin also attempts to draw the line between perception and introspection at the boundaries of the central nervous system:

“The physicalist distinguishes awareness (or discriminatory responses) characteristically caused by stimuli other than events in the central nervous system. Awareness of the second sort, or introspections, are characteristically caused by events of the first sort.”<sup>10</sup>

---

<sup>8</sup> Churchland 1988, p. 74.

<sup>9</sup> Churchland 1985, p. 16.

<sup>10</sup> Levin 1985, p. 131.

According to this story, introspection is a neural response caused by stimuli *in the central nervous system*, while perception is a response caused by stimuli somewhere in the peripheral nervous system.

In my opinion, this suggestion does not provide sufficient conditions for introspection. Not every neural event that is triggered by stimuli somewhere inside the CNS will count as a case of introspection. That would for instance entail that all motor signals created in the motor cortex are introspective responses in so far as they respond to activation in the sensory areas of cortex. And it would also entail that what we normally refer to as *visual perception* is in reality introspection, since the retinas are part of the central nervous system. The activation pattern on the retina are originally caused by events which are external to the CNS, but the neural signals leading from the retina to the visual cortex would, according to Levin's definition, be introspections since they are caused by stimuli from the retinas.<sup>11</sup>

Given a non-dualistic metaphysics it is hence rather difficult, given our current knowledge about the functions of the brain, to delimit introspection from perception solely by describing the objects of each process. For a Cartesian dualist, on the other hand, the object of introspection is partly defined in terms of what it does (it thinks and so on), and partly in terms of what kind of epistemic access we have to it. The Cartesian definition of "inner" is hence partly made by means of the special epistemic authority each person enjoys about this area of reality.

### 1.3 THE OBJECT OF INTROSPECTION IS THE SENSORY ASPECT OF MENTAL STATES

E. B. Titchener argued, just as his teacher Wilhelm Wundt<sup>12</sup> did, that the object of introspection consists in the *sensory* elements of consciousness. Sensations are according to Titchener the basic elements of the mind. They can be described in terms of different attributes or aspects. Titchener writes:

---

<sup>11</sup> It could be that Levin is out to demarcate the processes that "are about" other neural processes from the processes that "are about" things outside the nervous system. But, "be about" can not just like that be translated to "be caused by".

<sup>12</sup> Wundt 1999 [1897].

“Some sensations have four such aspects; every sensation has at least three. The four are quality, intensity, extent and duration. The process is itself, and not some other process (quality); it is stronger or weaker than other sensations (intensity); it spreads over a certain portion of space, greater or less (extent); and it lasts a certain, longer or shorter period of time (duration).”<sup>13</sup>

We do not however enjoy introspective awareness of *meanings*, or to what nowadays is usually referred to as “intentional content”, according to Titchener. The content of a perception or a thought is not a psychological phenomenon, Titchener argues, even if it can be “carried by all sorts of sensational and imaginal processes”.<sup>14</sup> The understanding of the meaning of a word is not made consciously and is therefore not a mental process since the mental or the mind is, in Titchener’s view, equal to the conscious mind. There are, however, “psychological vehicles of logical meanings” that can be introspected, but they should not be confused with the meaning or the intentional content associated with those vehicles. Titchener finds that even the concept of meaning *itself* can be ideated and its psychological vehicle introspected.

“Meaning in general is represented in my consciousness by another of these impressionist pictures. I see meaning as the blue-grey tip of a kind of scoop, which has a bit of yellow above it (probably part of a handle), and which is just digging into a dark mass of what appears to be plastic material.”<sup>15</sup>

He does however not think, which is his point, that the scoop *is* the meaning of meaning. The experience of it is just a psychological concomitant to the concept of meaning.

The task of psychology is hence to describe our experiences, but only in terms of their sensory properties, i.e., how clear or obscure, vivid or faint they are, and not at all in terms of what they mean or are about. To learn how to describe experiences in this way is a rather difficult task and that is why both Wundt and Titchener emphasised the need for *experienced research subjects*. A novice may find it hard to decide just how intense, obscure or

---

<sup>13</sup> Titchener 1998 [1896], p. 30.

<sup>14</sup> Titchener 1964b [1909], p. 181.

<sup>15</sup> Titchener 1964a [1909], p. 170.

vivid a certain experience is and he may almost certainly find it difficult not to describe an experience in terms of what it is an experience of. Should the subject all the same fail to do so, it is the experiment leader's task to omit all allusions to meaning from the protocol, according to Titchener and Wundt.

As long as one stays true to only describing the sensory aspects of an experience one is in a good position to avoid the mistake of sliding over from describing the mind to describing *the external world*, which was supposedly not the task of the psychological science. Psychological science was not supposed to concern itself with either the external world, (including the brain) or with the relations between the words and their meanings, but only with how any given mental event is experienced.<sup>16</sup>

William James argues that it is not the experiences in the "presently conscious" that constitute the possible objects of introspection. He thought it impossible to investigate our thoughts at the same time as we think them since that would entail a split of the conscious stream. What we really do (or can do) is examining our experiences after we have experienced them. It is only after we have experienced something that we can reflect upon what it was like to undergo the experience in question. The object of introspection will hence be the contents in "the short-term memory store", to put it in modern terms. James agreed with Titchener and Wundt that it is the *sensory aspects* that we introspect, but in his case it is the sensory qualities of "what just went by in our consciousness".

Titchener was a student of Wilhelm Wundt. Another of Wundt's students was Oswald Külpe. Külpe was the founder of a research programme, which aimed to investigate the thought processes. Külpe and his

---

<sup>16</sup> What if colours are properties of external objects? Then reports about colour patches are really not reports about the subject's own mind but about the external world. Humphrey (1951), e.g., analysed the reported introspections of Titchener's group at Cornell and came to the conclusion that the subjects in the experiments did not describe the sensory aspects of their experiences after all, but the properties of the external objects that they perceived. According to him the subjects of Titchener performed the exact error Titchener explicitly wanted to avoid, i.e. the "stimulus error". C. S. Peirce came to the same conclusion. He writes:

"In a certain sense, there is such a thing as introspection; but it consists in an interpretation of phenomena presenting themselves as external percepts. We first see blue and red things. It is quite a discovery when we find the eye has something to do with them, and a discovery still more recondite when we learn that there is an *ego* behind the eye, to which these qualities properly belong." (1950, p. 308.)

According to Peirce it is more probable that properties like redness, magnitude, intensity and so forth are properties of external objects than that they are properties of our experiences.

colleagues “found” something when they were investigating these processes.<sup>17</sup> The subjects often reported about an experience that occurred between the time they were presented with a stimulus and the time they gave their verbal reactions to this stimulus, an experience that was not at all like an image and hence not sensory in the way Titchener took all experiences to be. The Würzburg psychologists coined a new term for this phenomenon; *Bewusstseinslagen*, which has been translated as “imageless thoughts”. These, they argued, were non-sensory conscious processes, yet they were accessible by introspection.<sup>18</sup>

Is it Titchener’s opinion that a person cannot know what she is thinking about? – No, but he claims that she cannot know this by means of introspection. No one has first-person access to meanings, but the subject can, according to Titchener, “deduce” what she is thinking of from her introspective awareness of her experiences. It is not clear to me what kind of deductions this would entail. By means of introspection we can find out what sensory qualities a certain idea has, we can describe the idea in terms of intensity, clarity, vividness and so on, but how do we go from knowing these things about a certain experience to knowledge what it is an experience of? How can Titchener *deduce* from his experience of the blue-grey tip of the scoop that this experience is a concomitant to the concept of meaning?

#### 1.4 THE OBJECT OF INTROSPECTION IS THE INTENTIONAL CONTENT OF MENTAL STATES

It is quite generally agreed today that we have introspective access not only to the “sensory aspects” of our experiences but also to their intentional content. It is not only the case that we by introspecting can come to know what a certain experience feels like in terms of how intense it is and so forth, we have also introspective access to what it is we think about. According to a

---

<sup>17</sup> Külpe 1909.

<sup>18</sup> These alleged findings became the subject of a scientific dispute. Did the subject’s reports really reflect an underlying reality of non-sensational introspectible experiences? If they did, it seems as if Wundt and Titchener were wrong about their characterisation of the conscious mind as being reducible to sensory “atoms”. Titchener (1964 [1909], p 183) stated that he himself had tried really hard to come across *Bewusstseinslagen* by introspection, but had failed to do so. Given that he, as he himself claims, is “versed in introspection and sufficiently objective in purpose”, he claims that there are reasons to doubt the results from Würzburg.

strong version of the so-called *representational theory* of mind, experiences and what it like to have them can exhaustively be accounted for in terms of their intentional content. This means that what Titchener and his colleagues referred to as sensory atoms are according to representationalism not intrinsic properties of experiences, but part of the intentional content of those experiences.

Does this mean that perceptual experience does not have any relevant intrinsic properties which the subject can become aware of by means of introspection? Fred Dretske (1995, 1999), Sydney Shoemaker (1996, 2002) and Michael Tye (1995b, 2002) argue that the conclusion must be that introspective awareness is not awareness *of* our mental states. We can, however, come to know things about our mental states, not by attending to the experiences themselves, but by attending to what those states are representations of. Other philosophers like William Lycan (1996b) and Peter Carruthers (2000) maintain that we are in a sense aware *of* our perceptual experiences while introspecting them even if “what it’s like” to have the experiences is determined by their contents.

### 1.5 THE OBJECT OF INTROSPECTION IS BEHAVIOURAL DISPOSITIONS

The behavioural psychologists turned psychological science “upside down”. What was previously considered the chief method available to a psychologist was now abolished altogether. The introspective method was, at best, regarded as unreliable, and at worst, as non-existing. The stronger critique derived from the idea that the sensory atoms described by the experimental psychologists were mere fiction. If the designated objects of introspection prove not to exist, the special method for accessing these objects is not likely to exist either. The subject of psychological science now consisted of behaviour and behavioural dispositions.

Gilbert Ryle still finds some use for the term “introspection” though, even if he does not believe in an inner mind in the Cartesian sense, or a “ghost in the machine” as he calls it. I learn about my own mind in the same way as I learn about that of other people, he says. Ryle argues that the philosophers and psychologists who believe in a covert consciousness, only accessible by introspection, commit a version of the *category mistake*, where one tries to explain a concept in terms of one sort of metaphysical category

when one should, in fact, assign it to another. Analysing mental terms in terms of a Cartesian soul is one example of such a mistake. The correct category for most mental terms is, according to Ryle, behavioural dispositions. “Mind” “signifies my ability and proneness to do certain things and not some piece of personal apparatus without which I could or would not do them.”<sup>19</sup>

“The questions ‘What knowledge can a person get of the workings of his own mind?’, and ‘How does he get it?’ by their very wording suggest absurd answers. They suggest that, for a person to know that he is lazy, or has done a sum carefully, he must have taken a peep into a windowless chamber, illuminated by a very peculiar sort of light, and one to which only he has access.”<sup>20</sup>

In reality, we learn things about our own mind by reflecting upon our behaviour. The “things” we find out by doing this are the dispositions behind our different behavioural acts.

As it turns out, Ryle’s suggestion seems to work better for some mental phenomena than for others. While it is quite apt to explain how we come to form beliefs about e.g., our character traits, it is hard to see how we can learn anything about e.g., our day-dreams or vivid memories by studying our own behaviour. By noticing that I on a regular basis disregard the piles of dirty dishes in the sink and watch TV instead, I may come to the conclusion that I am lazy or that I don’t like doing the dishes. But what behaviour of mine can I study in order to learn something about my daydreams? Ryle describes how we sometimes “catch ourselves” daydreaming or humming a tune. Catching oneself daydreaming is what other philosophers have referred to as “introspection”, he says. First we engage in some kind of behaviour, which is called “day-dreaming”, and then we become involved in reflecting upon this behaviour. The reflection is, evidently, also a kind of behaviour but not inner in a metaphysical sense. He also explains that in order to check whether we have properly understood a mathematical calculation or a philosophical argument, we put ourselves to the same tests as we use for other people, i.e., we ask questions about how the argument goes or ask the person to perform a calculation on paper.

Ryle is consistently unwilling to give any explicit account of what exactly it is we catch or test. What is it that has been going on when we, e.g.,

---

<sup>19</sup> Ryle 1990 [1949], p. 161.

<sup>20</sup> Ibid., pp.161-162.

were daydreaming? He refers to “silent soliloquy” and “talking to oneself”<sup>21</sup> but remains vague on what that really comprises. B. F. Skinner (1965) as well as K. Lashley (1923) and J. Watson (1913) before him claim that what we call “thinking” is a repressed form of ordinary speech. In thinking we use the same muscles as when we talk to other people. They refer to this kind of speech as “reduced”, “truncated”, or “covert”. Ryle, however, does not commit himself to the view of the behavioural psychologists that thinking is a kind of sub-vocal activity in the larynx, or to a physicalist version of that thesis where muscular activity is replaced with activity in the brain.

## 1.6 SUMMARY OF CHAPTER ONE

Ideas about the object of introspection vary with the preferred theory of mind. For the philosophers who argue that experiences have intrinsic qualities, these qualities are the decided objects of introspection. For philosophers who do not believe in special intrinsic mental properties, the most natural choice for introspective objects will be either the brain, behaviour, or intentional contents. But regardless whether one prefers to think about the mind as a unique substance, as “nothing but the brain” or as overt or covert behaviour, the object of introspection is still considered to be what is, in some sense, in that mind.

For a proper understanding of how a particular philosopher conceives “introspection” it is however not enough to elucidate her or his view on the object of introspective awareness. We must also try to describe how this philosopher accounts for the epistemic relations between introspective beliefs and that which the beliefs are about.

---

<sup>21</sup> Ibid., p. 163.

## 2. The epistemic status of introspection

On the whole, we seem to know our own minds better than we know the minds of our fellow human beings. With more or less ease, we are ready to answer questions about our hopes for the future, our fears and desires, our feelings and beliefs. A much debated issue in the philosophy of introspection concerns the epistemic status of introspective beliefs. The question can be framed thus: in what sense (if any) is self-knowledge more certain than knowledge about the external world and knowledge about other people's minds? Different answers to this question have been advanced. Philosophers have been defending almost every possible stance, from one claiming that self-knowledge is infallible due to logical necessity to one claiming that introspection does not yield knowledge at all. Before looking closer at the different positions in this area, we will review the related question about whether introspection can and should be used as a scientific method in psychology.

### 2.1 IS INTROSPECTION A RELIABLE METHOD?

For the experimental psychologists at the beginning of the 20<sup>th</sup> century the question of the epistemic status of introspective reports was very important, since introspection was the chief scientific method of the psychological science. In the early 19<sup>th</sup> century Auguste Comte issued two challenges to the possibility of a psychological science based on introspection. First, he claimed that introspection of intellectual activities (as opposed to emotive ones) is impossible. It entails an impossible split of consciousness, since one part of consciousness must inspect another. He writes:

“[T]he thinking individual can not split himself in two, one part of which would think while the other would watch the former thinking. The organ observed and the organ observing being, in this case, identical, how could any such act of observation take place? This supposedly psychological method is therefore radically faulty in principle”<sup>22</sup>

---

<sup>22</sup> Comte 1830, lesson 1, p. 34.

This seems to be a materialist version of the idea that consciousness is indivisible, a view advocated by e.g., William James.<sup>23</sup> James claims that consciousness can be described as a “stream” of mental events. In this stream one experience succeeds another. No two or more experiences or mental events can occur at the same time. Introspection being a kind of observation seems to imply that two mental events do occur at the same time, the event introspected and the event introspecting. This is not possible since consciousness according to James by definition cannot be divided. Nor is it possible (or plausible) to think that one person has two separate streams of consciousnesses. Introspection, if it exists at all, can hence not be the simultaneous observing of consciousness.

Another assumed problem with using introspection as a psychological method, which Comte also poses, is that introspection might quite possibly *distort* that which is being introspected – the conscious experience. The received view (already acknowledged by Hume) was that this problem first and foremost affects introspection of *emotions* and not so much introspection of thoughts. Franz Brentano, e.g., writes:

“In observation, we direct our full attention to a phenomenon in order to apprehend it accurately. But with objects of inner perception this is absolutely impossible. This is especially clear with regard to certain mental phenomena such as anger. If someone is in a state in which he wants to observe his own anger raging within him, the anger must already be somewhat diminished, and so his original object of observation would have disappeared.”<sup>24</sup>

Some different solutions to this problem were put forward. James suggested that instead of studying the anger (e.g.) while being angry, one could study it afterwards. Then one would be able to “coolly scrutinize” these mental events. He called this *retrospection*. If we turn our attention to mental events which happened a few seconds ago, we avoid not only Brentano’s worry that we might change the object of introspection by attending to it, but also that of the impossible split of consciousness.

It is, however, debatable whether such a move really would solve the “distortion problem”. The introspection of the memory of being angry would perhaps give us an adequate idea of the memory of being angry, but it

---

<sup>23</sup> James 1950 [1890].

<sup>24</sup> Brentano 1995 [1874], pp. 29-30.

is not certain that this memory of being angry gives a correct idea of what *the anger was* like. Brentano is reluctant to use retrospection for this reason:

“As everyone knows, memory is, to a great extent, subject to illusion, while inner perception is infallible and does not admit of doubt.”<sup>25</sup>

Brentano offers another psychological method, that of *inner perception*, which is not to be confused with *inner observation*. Inner perception is different from inner observation in that it does not distort its objects. It is not so much like watching the mental event “straight in the face”, but more like throwing it a glance. In inner perception thus, we do not actively focus on a mental event, but wait until we come to notice it anyway (accidentally), which we will sooner or later.<sup>26</sup>

Wilhelm Wundt’s way to avoid Comte’s second objection, that of distortion, was to only use *trained subjects* in his experiments. It takes great skill and practice to observe and report about one’s mental states with accuracy, Wundt thought. Wundt is reputed for not having admitted data from observers with fewer than 10,000 trials of experience in introspective report.<sup>27</sup> Titchener endorsed the same view. He writes:

“The training of which I have spoken, as necessary to a systematic introspection, is essentially the same as the training necessary to reliable observation in physics or biology.”<sup>28</sup>

The problem of distortion discussed by these psychologists primarily concerns the possible distortion of the object of introspection which the process might inflict on it. One could also quite possibly be worried about the possibility that the introspective process would not accurately capture the nature of the introspective object – that introspection makes the mental object appear in a way that does not correspond to how it really is. As we have seen, the latter problem emerges as a central issue when one tries to solve the problem of distortion by introducing the concept of retrospection.

---

<sup>25</sup> Ibid., p. 35.

<sup>26</sup> This method will apparently pose some other problems. The psychological experiments will take much more time to conduct if the experiment leader must wait until his subjects involuntarily notice the impression of the stimulus just flashed before them.

<sup>27</sup> Boring 1953.

<sup>28</sup> Titchener 1912, p. 446.

James Watson also puts forward some serious critique against introspection:

“[W]e find [in introspectionism] as many analyses as there are individual psychologists. There is no element of control. [...] There has never been a discovery in subjective psychology; there has been only medieval speculation”<sup>29</sup>

The behaviourist critique is metaphysical as well as methodological. Watson and his colleagues argue that there is no such thing as an inner world which can only be accessed by a special method, and even if there were, we cannot build a science on reports which cannot be double-checked. Brentano acknowledged the methodological problem.

“Just as the object of observation is unique – a unique life of which, as we have seen, we can only observe part – so the observer himself is unique, and no one else is in a position to check his observations. For someone else can no more apprehend my mental phenomena through inner perception than I can those that belong to him. In this respect [...] the natural sciences appear to be in a much more favorable position than psychology.”<sup>30</sup>

Brentano also brings forward a slightly different problem. If each person has only access to her own mental states, how can we make generalisations from introspective findings?

“[I]f we are restricted in our observation to one single individual, what else can we conclude but that our view of mental phenomena is extremely incomplete. Will we not inevitably fall into the error of mistaking individual peculiarities for general characteristics?”<sup>31</sup>

Brentano adds that it is, after all, not the case that we are unable to substantiate introspective reports by supporting evidence since we have, at least, indirect knowledge of other people’s minds. “The phenomena of inner life usually express themselves, so to speak, i.e., they cause externally perceivable changes.”<sup>32</sup> One kind of “perceivable change” is the sounds a person tends to make. The psychologist has access to the subjects’ verbal

---

<sup>29</sup> Watson & MacDougall 1929, pp. 16-17.

<sup>30</sup> Brentano 1995 [1874], p. 37.

<sup>31</sup> Ibid.

<sup>32</sup> Ibid.

reports which are based on their inner perceptions, and these reports do with some reliability reflect the mental phenomena undergone by the person reporting about them. Brentano adds that we would not understand each other so well, if our mental lives did not share many common traits.

## 2.2 FIRST PERSON AUTHORITY

A quite common idea throughout the history of epistemology has been that the knowledge each person has of her own thoughts, feelings and perceptions is in a fundamental way different from the knowledge she has of other things including other people's minds. The difference primarily lies in the epistemic superiority of first person knowledge over other types of (empirical) knowledge. This alleged feature is commonly referred to as "first person authority" or "privileged access". That each person has privileged access to the character of her mental states might mean different things. Sometimes it means that the subject is the only one who can know anything about her own mental states; or alternatively, it might mean that the subject is usually in a better position to form true beliefs about her own mental state, than any outsider. A related claim is that the only certain knowledge we have is knowledge about our own minds. Some specifications of first person authority will now be outlined.

### 2.2.1 Infallibility

Some philosophers argue that a person's judgments or beliefs about her own mental states cannot be false, the possibility of error is excluded by necessity. Descartes writes:

"I am the same who feels, that is to say, who perceives certain things, as by the organs of sense, since in truth I see light, I hear noise, I feel heat. But it will be said that these phenomena are false and that I am dreaming. Let it be so; still it is at least quite certain that it seems to me that I see light, that I hear noise, and that I feel heat. That cannot be false."<sup>33</sup>

In one interpretation, thus, first person authority consists in *the infallibility of first-person judgments*. If P is such a judgement, it is not possible that S believes

---

<sup>33</sup> Descartes 1973 [1911], p. 153.

P while P is false. Does this immunity to error extend to all our first person beliefs or only to some of them? To begin with, false beliefs about one's own desires and motives seem to most people (ordinary ones as well as philosophers) commonplace. For instance, say that Petra's friend tells Petra that Petra's favourite horse colour is palomino. Petra comes to believe, however falsely, that she likes palomino coloured horses best. A common move made by philosophers to exclude these cases from the infallibility claim is to restrict the domain of infallible beliefs to include only those which are passed about *occurrent mental states*, in contrast to those who are about dispositional states. Petra's horse colour preferences might be described as a dispositional liking. If her friend said "I know what you are thinking right now, you think that you only like palomino horses", this would probably not had been accepted by Petra, if she was not in fact thinking exactly this at that particular moment.<sup>34</sup>

David Hume once wrote:

"[S]ince all actions and sensations of the mind are known to us by consciousness, they must necessarily appear in every particular what they are, and be what they appear."<sup>35</sup>

And Franz Brentano:

"The phenomena of inner perception are a different matter. They are true in themselves. As they appear to be, so they are in reality, a fact which are attested to by the evidence with which they are perceived."<sup>36</sup>

Husserl makes a similar point. He agrees with Brentano that inner perception is always and necessarily adequate, and in that sense different from outer perception. What is found by inner perception is "self-evident".

"Not only is it self-evident that I am: self-evidence also attaches to countless judgements of the form *I perceive this or that*, where I not merely

---

<sup>34</sup> Yet another case to which this infallibility (as well as other forms of authority) may not extend is beliefs about such things as proneness to react with aggression. It might be quite hard for a person to decide how she will react during periods of extreme stress, or during events she has not yet encountered, such as perhaps a major accident, the death of a close relative, or war.

<sup>35</sup> Hume 1978 [1711-1776], p. 190.

<sup>36</sup> Brentano 1995 [1874], pp. 19-20.

think, but am also self-evidently assured, that what I perceive is given as I think of it, that I apprehend the thing itself, and what it is - this pleasure, e.g., that fills me, this phantasm of the mind that float before me, etc.”<sup>37</sup>

“Appearance” is usually contrasted with “reality”. Things may appear in ways that are different from how they are. A coin lying on the table is *really* round, but may *appear* elliptic to an observer. That all bachelors are unmarried men *is true*, but may *appear false* to a confused thinker. Both Hume and Brentano express the idea that the distinction between reality and appearances is not applicable to occurrent mental states. These states are necessarily as they appear to the subject.

There are however, philosophers, who contest this view and claim that a person can be wrong about the *true nature* of her occurrent mental states, which entails that an external object can appear to a subject in a way it truly is, while the appearance, in contrast, fails to reveal its true nature to its owner. Phrased in this way, the proposal sounds contradictory: some object appears to a subject in a way that it does not appear. Philosophers who argue that “this” is possible, do not put it in these terms. Instead they say something like the following: the perceptual state accurately represents an external object, but the introspective state misrepresents the perception. Lycan brings forward an example based on a “fraternity initiation rite” where a blindfolded subject is told that a knife is about to be applied to his throat, and then believes that he feels pain when an ice cube is applied to the area. He is wrong about his own experience because it is *really* an experience of cold and not one of pain, which he introspects it to be, Lycan says.<sup>38</sup> The success of this argument is, I think, questionable. Could it not equally well be argued that although the *object of his perception* is a cold ice cube the student, nevertheless, *perceives pain*. It could hence be a case of mis-perception, and not one of mis-introspection. Just like when we mistake one object for another in visual or auditory perception, as a cow for a horse or a larch for a blackbird, we might mistake an ice cube for a razor blade. These are however not completely analogous cases, since cold and pain are experiences in different sense modalities, while the song of larch and the song of blackbird are experienced in the same modality. A better example would perhaps be when you press your eyeball with your finger and see a colour patch. What

---

<sup>37</sup> Husserl 2001 [1900/1901] II pp. 87-88.

<sup>38</sup> Lycan 1996.

you “ought” to experience is a touch sensation, what you in fact experience is a visual sensation.

It is difficult to draw any definite conclusions from these examples. At what “level” does the mistake reside? Am I mistaking a finger for a colour patch? Do I falsely believe that what caused my experience is light reflected from a surface while it really was a finger? (Or differently put: did I represent the finger situation wrongly?) or do I perceive a touch sensation caused by a finger and introspectively judge that it is a visual sensation? Lycan should maintain the latter interpretation if he wants to stay true to his initiation rite example.

The pain sense does not appear to be quite analogous to any of the other sense modalities. Hot and cold, red and soft are by some philosophers thought to be properties of objects - the ice cube is cold and the teapot is hot – while they, by others again, are regarded as intrinsic properties of experiences. Pain is, however, not normally considered to be a property of external objects. We sometimes talk of “painful episodes” or a “painful procedure”, but never of a painful cactus or a painful hot stove.<sup>39</sup> All this may, however, be a matter of how we use the terms and not of the metaphysical status of pain and red and so forth. Pain could very well have the same metaphysical status as e.g., red and sour. These phenomena may all be physical properties of objects, or dispositional properties of objects (or intrinsic properties of experiences). Our reasons for not talking about painful objects may have other grounds. Shoemaker writes:

“Consider Jonathan Bennett’s example of phenol-thio-uria, which tastes bitter to three-quarters of the population and is tasteless to the rest (see Bennett 1968). If as the result of selective breeding, or surgical tampering, it becomes tasteless to everyone, I say it has become tasteless. And if more drastic surgical tampering makes it taste sweet to everyone, I say it has become sweet. But I don’t think that if overnight massive surgery produces intrasubjective spectrum inversion in everyone, grass will have become red and daffodils will have become blue [...] It think that our color concepts are, for good reasons, more ‘objective’ than our concepts of flavors.”<sup>40</sup>

---

<sup>39</sup> Wittgenstein 1989 [1953] invites us to imagine that surfaces of things around us have patches that produce pain when we touch them. “In this case we should speak of pain-patches on the leaf of a particular plant just as at present we speak of red patches.” §312, p. 104. As it turns out, he was wrong. We do not speak of nettles as having pain-patches.

<sup>40</sup> Shoemaker 1996c p. 259.

But both red and bitter are still understood as properties of objects, Shoemaker says, while pain is not. By tradition, the perceiver enjoys extra authority here. If someone claims to have a hurting knee, we do not say that she is mistaken just because we cannot find anything wrong with her knee. The semantics here partly reflects our interests, Shoemaker says. Our strongest interest in pains is to get rid of them, not to properly locate them (in the knee or in the head).

The infallibility thesis is historically and theoretically associated with a dualistic theory of mind. Given the theory that mental states are identical to brain states, it seems hard to defend the claim that introspective knowledge is, in principle, infallible since in this case introspective knowledge is not so very different from any other kind of empirical knowledge. Armstrong, for instance, says that if introspection is incorrigible, materialism about mental states *must* be false, but it is not false and therefore introspection is corrigible.<sup>41</sup> He is thus not willing to give up a good metaphysical theory in a hurry. Introspection, according to him is a self-scanning process in the brain. And it must be possible for such process to yield false results, Armstrong says.

One could also argue that first person present tense reports about the phenomenal character of experience are infallible most of the time, or even always, but only in this world and not due to logical or metaphysical necessity. Lycan argues that it is metaphysically possible that an “internal monitor” fires without any proper cause, but that it very rarely happens.<sup>42</sup>

“[T]he inner-sense view predicts that it is possible for a person to be unveridically conscious or aware of a sensation that simply does not exist. You might introspect a sharp, severe pain when there is in fact no pain at all.”<sup>43</sup>

The reason that the internal monitor seldom misrepresents what happens on the first level of consciousness, in *this world*, according to Lycan, is that the monitor and that which is being monitored are spatially close to one another. “The first-order sectors they scan are immured right there in the brain with them, and there is little to threaten the informational connection.”<sup>44</sup> This

---

<sup>41</sup> Armstrong (1963) uses the terms “infallible” and “incorrigible” synonymously.

<sup>42</sup> Lycan pictures that introspections are conducted with the help of an internal monitor.

<sup>43</sup> Lycan 1996, p. 19.

<sup>44</sup> *Ibid.*, p. 20.

argument is weak, in my experience technical devices often malfunction even if their parts are intimately connected to each other.

Wittgenstein says something that may be interpreted as: the infallibility of first person judgments about sensations is a reason not to count these reports as conveying knowledge. While it is meaningful to say that other people know or doubt that I am in pain, it is not meaningful to say that I am in doubt with regard to my own pain.<sup>45</sup> The incentive for this claim is, I think, actually the same as what made Brentano and Hume conclude that self-knowledge is infallible. Namely the idea that we cannot distinguish how something appears from our awareness of it. Wittgenstein's point is probably that in order for someone to pass a *judgment* about something there must be at least a possibility that she could be wrong. C. I. Lewis also articulates this argument:

“Apprehensions of the given which such expressive statements formulate, are not judgments, and they are not here classed as knowledge, *because* they are not subject to any possible error.”<sup>46</sup>

### 2.2.2 Indubitability

A slightly weaker account of first person authority is captured by the claim that first-person judgments about occurrent mental states are *indubitable*. They are immune to any possible doubt by the person who is making them, as well as by others. It is assumed that one can have *no grounds* for doubting the truth of a given belief about one's own occurrent mental states.

Different philosophers have drawn different conclusions from this alleged fact. For instance, while Wittgenstein thinks that the impossibility of doubt is a liability rather than an asset, Descartes argues that the indubitability gives first person judgments the highest possible epistemic security, and Sir W. Hamilton writes:

“The facts of consciousness are to be considered in two points of view; either as evidencing their own ideal or phenomenal existence, or as evidencing the existence of something beyond them. A belief in the former is not identical to a belief in the latter. The one cannot, the other may possibly be refused [...] Now the reality of this, as a subjective datum – as an ideal phaenomenon, it is absolutely impossible to doubt without doubting the existence of consciousness, for consciousness is itself this

---

<sup>45</sup> Wittgenstein 1989 [1953], § 246.

<sup>46</sup> Lewis 1946, p. 182; italics added.

fact; and to doubt the existence of consciousness is absolutely impossible; for such a doubt could not exist, except in and through consciousness, it would, consequently, annihilate itself.”<sup>47</sup>

Descartes claims that we can be absolutely certain about our own existence, because it is impossible to doubt and not exist at the same time. But he also argues in the Meditations that follow, that while we can have doubts about the existence of external things, it is impossible to doubt, not only *that we are* conscious, but also *what we are conscious of*. I cannot (unless I know that God exists) be certain whether I am dreaming about an orange pumpkin or actually seeing one, but I cannot possibly doubt that I am, in some sense, aware of a pumpkin.

It all seems very plausible. How could anyone possibly doubt her own experiences? She cannot doubt that she has them; she cannot doubt that they are hers and she cannot doubt that they are the way they are – what they feel like to her. A different angle on this problem is captured by the following: “sure, you can have no reason to doubt your own opinions about your experiences, but this indubitability does not bestow these opinions with some special epistemic certainty, it entails that the opinions have nothing to do with knowledge at all.” In other words: that we can have no grounds for doubting our judgments about sensations is another way of saying that they are not *justified*. Crispin Wright says in the spirit of Wittgenstein that “phenomenal avowals” such as “I have a headache”, are groundless.<sup>48</sup> There is simply nothing one can say to substantiate one’s claim besides “I know that I have a headache because I feel it.”

It might be the case that the certainty that is intimately connected with first person judgments is not epistemic at all, but merely a psychological or phenomenological fact. Being certain about something could be understood as a matter of *feeling* assured or confident that one is correct. Certainty construed as a psychological impossibility to doubt one’s own first person beliefs is of course different from the logical impossibility of doubt or the “semantic” impossibility, discussed by Wright.

Descartes claimed that “clarity” gives epistemic certainty. If a certain idea stands clear before my mind, I am entitled to judge that the idea is true. After establishing the certainty of his own existence as well as that of God by means of deductive arguments, Descartes turns to reflecting upon what it’s

---

<sup>47</sup> Hamilton 1874, p. 188.

<sup>48</sup> Wright 1998, p. 14.

like to entertain these ideas that he already knows are true. He notices that these true ideas (the idea of an existing God, and the idea of a self) stand “clearer before his mind” than other ideas do. He goes on to use the clarity which pertains to these ideas as a criterion of their truth.

“[It] seems to me that already I can establish as a general rule that all things which I perceive very clearly and very distinctly are true.”<sup>49</sup>

It is not clear whether Descartes uses “clarity” in an epistemic or in a phenomenological sense (or both) in his *Meditations*. But I think it is fair to say that he at least partly means clear in a phenomenal sense.

“I term that clear which is present and apparent to an attentive mind, in the same way as we assert that we see objects clearly when, being present to the regarding eye they operate upon it with sufficient strength. But the distinct is that which is so precise and different from other objects that it contains within itself nothing but what is clear.”<sup>50</sup>

Descartes seems to think that clarity is an experienced, phenomenal fact and that its presence works as justification for an affirmative judgment about the idea to which this clarity pertains. If I perceive the existence of God clearly, I can be absolutely certain that God exists, and the clarity of the idea justifies my belief in the existence of God.

### 2.2.3 Incorrigeability by another party

Some philosophers have settled for an account of first person authority, which entails that even though it is possible for *me* to doubt and correct my own first person beliefs, my judgments cannot be corrected *by others*. This has been referred to as *incorrigeability of first-person judgments*.<sup>51</sup> If I claim that something appears red to me, I might possibly doubt the correctness of this judgment, but you cannot show that I am mistaken (and therefore you can have no reasons to doubt whether I am right or not).

---

<sup>49</sup> Descartes 1973 [1911], p. 158.

<sup>50</sup> *Ibid.*, p. 237.

<sup>51</sup> Alston 1989, e.g., uses the term “incorrigeability” to refer to the assumed impossibility of another person to correct someone’s first person reports. But incorrigeability in a wider sense claims that no one including the person herself can correct a person’s reports about her own mental states. Incorrigeability in this wide sense seems to coincide with indubitability.

Ayer formulates the notion of incorrigibility thus:

“If this is correct, it provides us with a satisfactory model for the logic of the statements that a person may make about his present thoughts and feelings. He may not be infallible, but still his word is sovereign. The logic of these statements that a person makes about himself is such that if others were to contradict him we should not be entitled to say that they were right so long as he honestly maintained his stand against them.”<sup>52</sup>

I could, hence, be wrong in thinking that I seem to be seeing something green now, but if I am wrong, *correction can only come from me*.

Armstrong challenges this incorrigibility thesis. He invites us to imagine a “brain super-technician” who can check whether my reports about pains and the like are supported by what he can see when he is examining the brain. It might be objected, Armstrong admits, that there is no reason why we should side with the brain technician. If he and I disagree, is it not possible that there is something wrong with the theory of how mental states and brain states correlate? Armstrong does not really have to show that we *must* side with the brain technician every time he and the person having an experience disagree. He must only show that it is *possible* that the technician might be right, and hence that the introspective reports given by a person is, in principle, open to correction from others.

Sometimes, we do seem to take *behavioural evidence* over introspective reports, Armstrong maintains and gives us an example of when we, in fact, tend to side with the other party than the introspector in a dispute over his judgments about his own sensations.

“I say perfectly sincerely, ‘I am in great pain.’ A little later I inquire why nobody gave me any assistance or sympathy when I shrieked out ‘I am in great pain’ and exhibited every sign of distress. It is then proved to me that I said the words in a quite, level voice while exhibiting every sign of relaxation. Might it not be reasonable to conclude that I was mistaken in thinking myself to be in pain?”<sup>53</sup>

Robert Audi makes the same point by presenting the following example: imagine a fourteen-year-old boy who has a fear of heights. When he is asked to come and climb a mountain with his fellow campers he declines with the

---

<sup>52</sup> Ayer 1963, p. 73.

<sup>53</sup> Armstrong 1963, p. 425.

explanation that his knee hurts. Suppose that he is excused from the hike and immediately after the hikers leave he goes and plays tennis without visible difficulties. When asked whether his knee hurts, he says that it does but that he is playing in spite of this. One might say that the boy's enthusiasm for tennis has overshadowed the pain, which is still present, but this is not likely if the severity of the pain made the boy limp a moment ago, Audi says. Suppose that we know that the boy hates to entertain thoughts claiming that he is afraid of doing things boys are supposed to do with ease, *and* that he believes that carrying on despite pain is a mark of true masculinity, which he is in favour of. We then have a good understanding of why it would be comforting for the boy to believe that he has a pain in the knee. We also have medical and behavioural evidence that he does not have a pain. Is it not possible that the boy actually believes that he has a pain in his knee when he does not? Audi argues that it is possible both that the boy's belief that he is in pain is false and that we could have overriding reasons to say that it is false.

“The plausibility of this conclusion is strengthened when we reflect on the not uncommon practice of regarding with incredulity certain complaints of pain by neurotic invalids who crave attention, or spoiled children who are being ignored. The complaints in question are not regarded as insincere, particularly since in such cases there is often some slight discomfort; but when such people's behavior fails to cohere with the supposition that they are in pain and is at the same time readily interpretable as meant, however subliminally, to get attention, it seems quite proper to disbelieve at least some of their perfectly sincere complaints of pain.”<sup>54</sup>

#### 2.2.4 Transparency

An idea that used to be quite generally held was that a subject knows everything there is to know about her own mind. In other words, it is not only the case that we can be certain of our first person beliefs, we can also be certain that we know *everything* about our own mental states - nothing escapes the introspective eye. This feature has been called “omniscience”. We are so to speak “immune to ignorance” when it comes to our own mental states.

This assumption made some sense when one also claimed that the mind equals the conscious mind. According to Descartes the mind is in its essence self-transparent; we cannot be conscious of something and at the

---

<sup>54</sup> Audi 1993, p. 172.

same time not know about us being thus conscious. Titchener e.g., held the same view: there is no such thing as a process or event being both unconscious and mental. A mental process can be reduced to sensations, which by their very nature reveal themselves to the subject. One cannot experience something unknowingly.

There are also contemporary philosophers who argue for a limited version of the omniscience thesis that might be called the “transparency thesis”. Shoemaker e.g., claims that it is impossible for a *rational agent*<sup>55</sup> to believe something and not at the same time believe that she believes it. Denying the possibility of self-blindness in this sense, (a self-blind creature would be one that could entertain beliefs and be unable to self-ascribe those beliefs) does not mean that one has to reject the existence of unconscious beliefs and desires. To unconsciously believe that one is a failure, e.g., does not entail that one is ready to assert: “I am a failure but I do not believe that I am a failure.” It is the latter belief that a rational agent cannot possibly entertain.

“What I have asserted, in denying the possibility of self-blindness, is a connection between self-knowledge and rationality; that given certain conceptual capacities, rationality necessarily goes with self-knowledge. It is entirely compatible with this that there are failures of rationality that manifest themselves in failures of self-knowledge.”<sup>56</sup>

The “immunity to ignorance” that we enjoy according to Shoemaker also entails that it is impossible to feel something while not being aware of it.

It is widely accepted among philosophers that if a certain proposition, P ascribes to a person a certain phenomenal state like “being in pain” the person in question could not be ignorant of the truth of P. It is in other words not possible to be in pain and not know about it. There are, however, more moderate versions of this thesis as well. Wright e.g., writes:

---

<sup>55</sup> Shoemaker’s “rational agent” is an agent who, if she believes that P, will use P as a premise in her reasonings. The rational agent will also know “that to act on the assumption that something is true is to act as if one believes the thing; and she will know that if it is in one’s interest to act in this way it will normally be in one’s interest to make manifest to others that one is so acting – this will increase the likelihood that other believers in the truth of the proposition will cooperate with her in endeavors whose success depends on the truth of the proposition, and it will tend to promote belief in that proposition within her community and so to promote the success of endeavors whose success depends both on the proposition being true and on its being believed to be true by participants in the endeavors.” (1996, p. 82)

<sup>56</sup> Shoemaker 1996d, p. 49.

“[P]henomenal avowals exhibit a kind of *transparency*. Where P is an avowal of the type concerned, there is typically something absurd about a profession of the form, ‘I don’t know whether P’ – don’t know whether I have a headache, for instance, or whether my feet are sore. Not always: there are contexts in which I might be uncertain of a precondition – for instance, whether I have feet. But in the normal run of cases, the subject’s ignorance of the truth or falsity of an avowal of this kind is not, it seems, an option.”<sup>57</sup>

### 2.2.5 Immediacy

The notion of immediacy is connected to those of infallibility, incorrigibility, and indubitability. Self-knowledge is often said to be immediate, in contrast to knowledge of other people and the world around, which is indirect. What does “immediate” or “direct” mean in these circumstances? Norman Malcolm, e.g., claims that those who advocate a theory of direct perception hold that “direct” entails the *impossibility of error*. He constructs the following definition:

“A *directly* perceives *x* if and only if A’s assertion that he perceives *x* could not be mistaken.”<sup>58</sup>

Immediacy is thus simply analysed in terms of infallibility here. Conversely, it is a common assumption that we can only account for errors or mistakes in perception if we assume that the perceptual process is *indirect*. According to Hermann von Helmholtz and Richard Gregory perception is indirect because it involves *judgments* about sensory information.<sup>59</sup> An unveridical perception is a false belief about the world and beliefs about the world are based on the sensory information we receive. If I see a ghost where there is no ghost, I have interpreted sensory information wrongly which has led me to pass a judgment that is false. “Indirect” thus means “inferential” here.

Another psychologist, James Gibson, argues that the objects (surfaces) in the world around us are *directly* perceived. He claims that our access to external objects is not mediated by perceptual representations, sense data and the like.

---

<sup>57</sup> Wright 1998, p. 15.

<sup>58</sup> Malcolm 1963, p. 89.

<sup>59</sup> von Helmholtz 1924-1925, Gregory 1980.

“Direct perception is what one gets from seeing Niagara Falls, say, as distinguished from seeing a picture of it. The latter kind of perception is *mediated*. So when I assert that perception of the environment is direct, I mean that it is not mediated by *retinal* pictures, *neural* pictures, or *mental* pictures.”<sup>60</sup>

According to Gregory and Helmholtz perception is not only indirect in the sense that it is logically inferential, they also hold that it is also causally mediated. In Gibson’s opinion, perception is without doubt non-inferential, but it is not entirely unmediated. Critics of Gibson’s theory have argued that one of the major flaws of Gibson’s theory is that it is inapt to account for perceptual errors. Some of the perceptual mistakes we make cannot be explained if we do not assume that perceptions are based on inferences from sensory data, the critics argue. If I see a ghost in the room when I am looking at what is really a coat on a chair, the most plausible explanation seems to be that I have misinterpreted the sensory information somehow. And how can Gibson explain the perception of the famous duck-rabbit? When we shift from seeing the rabbit to seeing the duck and vice versa, nothing in the picture has changed, and if we stay focused on the picture, nothing “inside us” has changed except our interpretation of the stimulus, the inferentialists argue.

Gibson does not want to evoke a theory of epistemic indirectness, but is willing to take a step towards causal mediacy. The reason we sometimes see things that do not exist is because physical and/or neurophysiological factors have gone astray. For example, Gibson mentions that mirages and curved oars in water can be explained by the fact that light travels through a non-homogenous medium and is therefore diffracted in an unusual way. This means that in one sense, direct perception is mediated, since we see physical objects *via light*.

“Direct” and “indirect” are, thus, used in two different senses here, which can be described as an *epistemic sense* and a *metaphysical sense*, respectively.

(i) Epistemic directness signifies that we can grasp the objects of knowledge, (in this case our mental states) directly without having to rely on any other knowledge we have.

---

<sup>60</sup> Gibson 1979, p. 147.

(ii) Metaphysical directness claims that there are no causal and/or representational intermediaries between my belief in something and the object of this belief.<sup>61</sup>

What does “not having to rely on knowledge of something else” in order to know something really mean? G. E. Moore has in a series of books and articles tried to account for what is referred to as “immediate knowledge”. In *Some Main Problems of Philosophy* he characterises “direct apprehension” as a relation a subject has to a sense datum when he feels it or sees it. Propositions can, however, also be directly apprehended and must be so in order for the subject to understand the proposition. By directly apprehending a proposition which is about something, we indirectly apprehend that which it is about. If I e.g., believe that the moon is a yellow cheese, I directly apprehend the proposition “the moon is a yellow cheese” and indirectly, the moon. “Immediate knowledge”, then, is described as follows:

“If, therefore, either my argument or my opponent’s, or any other argument whatever, is to be a good one, it must be the case that we are capable of knowing at least *one* proposition to be true, *without* knowing any other proposition whatever from which it follows. And I propose to call this way of knowing a proposition to be true, *immediate* knowledge.”<sup>62</sup>

To know that a certain proposition is true without knowing any other proposition from which it follows, seems to be to know the truth of the proposition non-inferentially. I have not inferred the truth of the proposition in question from any other propositions.<sup>63</sup>

It is important to note that one primary motive behind distinguishing between “direct” and “indirect knowledge”, in the first place, was to give an account of how knowledge of our own minds differs from knowledge of the physical world (and of other people’s minds). C. I. Lewis writes:

---

<sup>61</sup> For the representationalist version of (ii), see below.

<sup>62</sup> Moore 1953 [1873-1958], p. 123.

<sup>63</sup> Helge Malmgren (1983) argues that Moore is actually working with two connected concepts of immediate knowledge. The first, and stronger one, is the one accounted for above. The second, slightly weaker one is counterfactual and says that you know p immediately if no other knowledge from which p follows is in this situation necessary for knowing p. The weaker concept entails, thus, that I know p immediately if my knowledge of p does not presuppose knowledge of something else.

“It is one such essential feature of what the word ‘mind’ means that minds are private; that one’s own mind is something with which one is directly acquainted – nothing more so – but that the mind of another is something which one is unable directly to inspect.”<sup>64</sup>

In what sense is introspection taken to be direct? According to the standard characterisation of introspection of experiences, we do not base judgments about our phenomenal experiences on inferences. We do not infer “I feel pain now” from any other propositions. But, according to e.g., Armstrong, an introspective belief can be doubted if behavioural evidence speaks against it. If I seem to be feeling pain, but am at the same time behaving quite carefreely I have reason to believe that my judgment about pain is wrong. I use the collateral evidence of not wincing, or moaning, or staggering to disprove my belief that I am in pain.

According to Ryle and the behavioural psychologists, introspective knowledge is inferential or it at least presupposes knowledge of something else, namely our behaviour. A typical introspective judgment is, in Ryle’s view, based on observations of our own behaviour. A typical introspective judgment is however, not “I am in pain”, but rather “I am a lazy student”, according to him.<sup>65</sup>

Is introspection mediated in some other sense? According to William Alston, it is pointless to discuss metaphysical causal directness when it comes to introspection, since we do not know how to determine what that would entail. “We are not able to assign precise spatial locations to mental states.” Alston continues:

“But if I am to determine whether my desire to go to Europe is an immediate causal antecedent of my belief that I have a desire to go to Europe, I need to have a more fine-grained view of the causal processes involved, and unfortunately we do not have any such view.”<sup>66</sup>

---

<sup>64</sup> Lewis 1964, p. 332.

<sup>65</sup> Donald Hebb thinks that the inferentiality of first person knowledge is indeed a reason to reject introspection as such. He writes: “All available information points to the conclusion that introspection does not exist; knowledge of mind is inferential, not phenomenal.” 1977, p 213.

<sup>66</sup> Alston 1989, p. 267.

However, metaphysical mediacy need not necessarily be *causal*. Perceptual awareness may for instance be mediated by sense data. According to Armstrong and Lycan inner awareness is caused by those perceptual states it is about, but it is also mediated in a different sense. Perceptual awareness of physical objects is, in their opinion, mediated by mental representations that represent the physical objects, and introspective awareness is mediated in a similar sense. If a subject is introspectively aware of a perceptual experience (which represents e.g., a brown horse), she is aware of her experience via a representation of the experience.

Let us recapitulate the different interpretations of “first person authority” made so far. Assume that P is a proposition that ascribes a phenomenal state to a person (S). The different epistemological claims discussed above can then be described as follows:

T1. *Infallibility*: It is not possible for S to believe that P is true while P is false.

T2. *Indubitability*: If S believes P there are no rational grounds for S for doubting that she knows that P is true.

T3. *Incorrigibility (by another party)*: It is not possible for S to believe P and for someone else to show that S is wrong.

T4. *Transparency*: It is impossible for S to remain ignorant of P if P is true.

T5. *Epistemic Immediacy*: S’s knowledge of P is not inferred from any other propositions, nor does it presuppose knowledge of anything else.

T6. *Metaphysical Immediacy*: S’s awareness of P is not mediated, causally or representationally, by anything.

### **2.2.6 Are first person beliefs justified?**

Below, I will discuss some related problems that concern the epistemic status of introspective beliefs. As you might remember, it has been argued that first person beliefs are not justified. This conclusion is sometimes inferred from the fact that such beliefs are impossible to doubt, and at other times it is taken to be a consequence of them not being based on some other knowledge. The traditional analysis of “S knows P” looks like this:

A.S believes that P.

B.S is justified in believing that P.

C.It is the case that P.

The infallibility thesis holds that if P is an introspected fact, (C) is true if (A) is true. The indubitability thesis seems to be advocating that B is true if A is true, but that depends on whether we want to analyse “having no grounds for doubt” as a kind of justification or not. One possible interpretation is that S is justified in believing P if (and only if) S does not have any grounds for doubting that S knows P. A different point which is argued by Wittgenstein (in one interpretation of him) and also by Shoemaker is that if it is correct that S has no grounds whatsoever for doubting P, she cannot be justified in believing P, since justification entails the possibility of doubt.

What gives first person beliefs the status of knowledge? The definition of epistemic immediacy excludes the possibility of a justification or a “base” in other known propositions which could constitute this distinguishing mark. The subject is unable to provide any reasons or any evidence to substantiate her first person beliefs. Moore’s view on this matter seems to be, according to Malmgren that we do not *need* any “reasons” or “grounds” in the sense of supporting knowledge of other proposition in order to *know* things immediately. Moore writes:

“In the case of ‘I’ve got a pain’ it doesn’t seem right to say ‘I have *evidence* that I’ve got a pain’ ...if you’re asked: What evidence have you got that you’ve got a pain? you must answer. None. You can also say ‘None is needed: I know it without needing any evidence’.”<sup>67</sup>

On the other hand, Malmgren continues, immediate knowledge that I am in pain is in one sense based on the fact that I have a pain. Moore writes:

“It’s not because of my knowing or observing something else that I know I’m in pain, but because I’m in pain or rather because I have just that pain. [...] If I am in pain, & say that I am, I know that what I say is true, because I have a particular sort of pain.”<sup>68</sup>

---

<sup>67</sup> Moore 1962, p. 173.

<sup>68</sup> Ibid.

Is it then the mere presence of pain that is the justification for a subject's belief that she is in pain? Alston points out that both Ayer and Shoemaker seem to hold the thesis that justification of first person beliefs does come from the mere presence of phenomenal experiences. They both seem to argue that C implies B here, i.e., the fact that S is in pain justifies her belief that she is. Ayer writes:

“In other cases where knowledge is claimed, it is not sufficient that one be able to give a true report of what one claims to know: it is necessary also that the claim be authorized, and this is done by adducing some other statement which in some way supports the statement for which the claim is made. But in this case no such authority is needed ... Our knowledge of our thoughts and feelings accrues to us automatically in the sense that having them puts us in a position and gives us the authority to report them.”<sup>69</sup>

And Shoemaker:

“[I]t is characteristic of a certain kind of statements, what I there called ‘first-person experience statements’, that being entitled to assert such a statement does not consist in having established that the statement is true, i.e., in having good evidence that it is true or having observed that it is true, but consists simply in the statement’s *being* true.”<sup>70</sup>

That which justifies the belief is, hence, the truth of the proposition it expresses. This means that whenever I have a true belief about my own mental states, I know it. Alston calls cases of knowledge where nothing is required to satisfy B over and above the other conditions of knowledge – “autonomous knowledge”. This makes B, in a way redundant. We do not need to justify our first person beliefs, they still constitute knowledge. Autonomous knowledge does however not coincide with what is usually taken to be epistemically immediate knowledge, Alston argues. A direct realist about perception will hardly think that the truth of a proposition describing a state of affair in the world justifies a perceptual belief in this proposition.

In the light of these considerations, Alston proposes another possible interpretation of what first person authority might entail. One could take A instead of C as a sufficient condition for B, something he calls “self-

---

<sup>69</sup> Ayer 1963, p. 64.

<sup>70</sup> Shoemaker 1963, p. 216.

warrant". This entails that each person is so related to propositions ascribing phenomenal mental states to himself that it is impossible both for him to believe that such a proposition is true and not be justified in holding this belief. So in this case it is the belief in a proposition and not the truth of the same proposition that suffices for justification. Self-warrant can be regarded as a weaker analogue of infallibility since it leaves open the possibility of error. Following Alston we can add:

T7. *Autonomy*: If P is true, then S is justified in believing P.

T8. *Self-warrant*: If S believes P, then she is justified in believing P.

### 2.3 FIRST PERSON AUTHORITY AND EXTERNALISM

Externalism about mental content has been put forward as a challenge for first person authority. Externalism about content, in short, claims that the content of our thoughts is (partly) individuated by factors that are external to us. Mental content does not supervene on the intrinsic properties of a person. Externalism implies that it is possible for two persons who are alike in all intrinsic aspects to differ in the contents of their thoughts in virtue of differences in their environment. Two different externalist versions are often discussed in these circumstances, one that is motivated by the Putnamian Twin Earth examples concerning natural kinds<sup>71</sup>, and one that is defended by Tyler Burge<sup>72</sup> and which tends to show that social institutions play a role in determining the contents of some beliefs, including some that do not involve natural kind concepts.

Boghossian introduces the "slow switching thought experiment"<sup>73</sup>. Oscar has without his knowledge been travelling back and forth between Earth and Twin Earth. Twin Earth is the well-known planet that is just like Earth with just one exception, the liquid (twater) that fills the oceans, lakes and rivers is not H<sub>2</sub>O but of a different chemical composition, called X<sub>Y</sub>Z. This difference can only be detected on a chemical level; it is

---

<sup>71</sup> Putnam 1975.

<sup>72</sup> Burge 1979, 1986.

<sup>73</sup> Boghossian 1989.

indistinguishable from water in all other respects. With each move Oscar stays long enough on each planet to acquire the concepts of the locals. When Oscar, after some time on Twin Earth entertains the proposition “water is liquid” he does not mean the same thing on Earth and Twin Earth. On the latter planet he is really expressing the thought that twater is liquid, but he does not know that, since he thinks he is on Earth all the time.

In the most common interpretation, the first person authority thesis entails that we can know things about our own mental states without making empirical observations involving our sense organs. Each person is able to find out without such empirical investigations both a) that she is entertaining a thought, b) that it has a particular content, and c) what that content is (e.g., that water is wet). If we combine the first person authority thesis (FPA) with that of externalism (E) we seem to end up either in a position where we know certain facts about the world without conducting any empirical investigations, facts that it seems highly unlikely that we could learn simply by examining what is in one’s own mind (like the fact that H<sub>2</sub>O is liquid) *or* in the position where we do *not* know certain other things a priori, things that are commonly thought to be known a priori (such as “I am thinking about water now”).<sup>74</sup> There are basically three ways to avoid this conclusion, (i) reject E, (ii) reject FPA, (iii) argue that FPA and E can be combined but that the conclusion made above can be avoided.

McKinsey (1991), Brown (1995) and Boghossian (1998) all argue that the tension between externalism and first person authority is genuine and that semantic externalism leads to implausible conclusions about what can be known *a priori*. If externalism is correct, thoughts with wide content depend on the presence of certain substances (or social practices) in the environment. If we have privileged knowledge of our thoughts, we would know a priori that our environment contains a certain natural kind and that seems absurd.

Other philosophers have argued against this incompatibilistic conclusion. Tyler Burge claims that E and FPA become compatible if we

---

<sup>74</sup> The term “a priori” is often used in these discussions. “Knowable a priori” should here be understood as “knowable without conducting any empirical investigations which involve the sense organs”. Introspective self-knowledge is then by definition taken to be a priori. One could, of course, also look upon introspective knowledge as knowledge a posteriori, especially if one thinks that the introspective process is perception-like. Then if we enjoy first person privileges and if content externalism is true it does not follow that we know things about the world a priori, just that our beliefs about the world are not based on sense experiences (which seems to pose just as big a problem).

maintain that in order to know the content of one's thoughts it is not required that one knows the environmental conditions that makes such thoughts possible. Even if externalism is true, you cannot be wrong in your beliefs about what you believe since what you believe you believe is entailed by what you believe. If you believe that there is water in the pond, and introspectively believe that you believe that there is water in the pond, whatever externalist semantics makes the first of those mental tokens of "water" designate H<sub>2</sub>O also makes the second of them designate the same substance. Introspective beliefs do not need any extra justification.

"In basic self-knowledge, one simultaneously thinks through a first-order thought (that water is a liquid) and thinks about it as one's own. The content of the first-order (contained) thought is fixed by nonindividualistic background conditions. And by its reflexive, self-referential character, the second-order judgment is logically locked (self-referentially) onto the first-order content which it both contains and takes as its subject matter."<sup>75</sup>

Dretske proposes that we should hold on to the metaphysics and re-examine "a suspect epistemology". Dretske claims that there is something wrong with the whole incompatibilism argument, since it assumes that knowledge of the content of your mental states requires knowledge that you have a mind. This is false, he says, it is quite possible to know what you think about and not at the same time know that you think it.

"It assumes that knowledge of what you think – for instance, that there is water – is (or requires) knowledge that you think. This is false. The special authority we enjoy about our minds is an authority about what we think – that, for instance, there is water – not about the fact that we think it."<sup>76</sup>

Dretske makes a comparison with the representing powers of instruments. Any simple measuring instrument can be construed to carry information about how it represents the world as being, Dretske says. Then we can modify the instrument so that it not only says something about external affairs, but also about how it is representing these external affairs.

"We can give it this additional function by simply affixing an additional label to the scale. Originally it was labeled 'Value of *Q*.' We now add the

---

<sup>75</sup> Burge 1988, pp. 659-660.

<sup>76</sup> Dretske 2003b, p.133.

label ‘Value that  $Q$  is represented as having.’ Now, when the pointer points at ‘5,’ the instrument does two things: it represents some external object as having a  $Q$  of 5 and it represents itself as representing this object as having a  $Q$  of 5. [...] [T]he instrument does two representational jobs by occupying one physical state. The instrument is fallible about the first thing it says, but infallible about the second.”<sup>77</sup>

The same things hold about us human beings according to Dretske.

Dretske’s argument shares some similarities with the one given by Burge. Burge claims that it is a mistake to think that that either perceptual knowledge or introspective knowledge require as a *precondition* knowledge of how we can know these things. He writes:

“It is a fundamental mistake to think that perceptual knowledge of physical entities requires, as a precondition, knowledge of the conditions that make such knowledge possible. Our epistemic right to our perceptual judgments does not rest on some prior justified belief that certain enabling conditions are satisfied.”<sup>78</sup>

Dretske says that although sense perception can give reliable information about what is in the world, it cannot give information about there *being a world*.

“To know that there is an external world, you have to discover it in some way other than by sense perception – the faculty that, if there is such a world, tells you what is in it.”<sup>79</sup>

So, given that the proper external relations obtain, the content of the subject’s thoughts and experiences are infallibly internally accessible to the subject. In the McKinsey-Boghossian argument for incompatibilism it is assumed that the subject *knows* that semantic externalism is true. The truth of externalism is according to Dretske and Burge not given by either perception or introspection. As little as I know that I am thinking about H<sub>2</sub>O, when I am thinking about water, do I know that water is H<sub>2</sub>O when I think “I believe that I think about water.”

These issues have been much debated. A number of different articles have recently been published which aim to show in different ways that first

---

<sup>77</sup> Ibid., p. 134.

<sup>78</sup> Burge 1988, p. 654.

<sup>79</sup> Dretske 2003b, p. 137.

person authority and semantic externalism are not incompatible. (See e.g., the papers in Nuccetelli 2003). This is however not the place and time to go through them all. It suffices to say that it seems to many philosophers that both semantic externalism and the thesis that self-knowledge is privileged are plausible theories (for different reasons), but combined they pose epistemic problems. Problems that according to some philosophers are unsolvable and therefore either externalism or first person authority need to be rejected. Others again, like Dretske and Burge, think that these problems can be worked around.

### 2.3.1 Phenomenal externalism

The Twin Earth example does not, however, seem to affect beliefs about the phenomenal character of mental states, which is the primary concern of the first person authority thesis. It does not matter whether we are on Twin Earth or Earth; the substance in the oceans is still perceived as being wet, cold and salty-tasting. If we argue that infallibility of first person beliefs only extends to beliefs about what phenomenal properties we experience, the problem which externalism posed seems to disappear. The belief that something appears red to me right now cannot possibly be wrong. That, however, depends on whether we think that the contents of our perceptual beliefs about how things appear are determined by the environment or not.

Some philosophers maintain that the phenomenal character of perceptual experiences is part of their representational content and that this content is “wide”, just as the content of beliefs nowadays usually is taken to be. Lycan, Tye and Dretske advocate such a view.<sup>80</sup> This entails that phenomenal character does not supervene on anything inside the subject’s head. Two persons could be internally exactly alike in every detail, but still experience the world differently. Ned Block e.g., argues that “phenomenal externalism” in the sense outlined above, is inconceivable. Block presents another imaginary planet, one which he calls “Inverted Earth” to show this.<sup>81</sup>

Inverted Earth differs from Earth in only two respects: everything has the complementary colour of the colour it has on Earth, and the colour vocabulary is also inverted. So, on Inverted Earth the sky is yellow and the

---

<sup>80</sup> Dretske 1996 labelled the position “phenomenal externalism”.

<sup>81</sup> Block, 1990. The inverted spectrum scenario was first put forward by Locke. In *Essay* Locke entertained the possibility that a Violet might produce in one man’s mind the same that a Marigold produces in another man’s and vice versa. Block’s argument is an inverted version of spectrum inversion.

grass is red and if you ask some of the inhabitants what colour the sky has, he or she would answer “blue“. Block emphasises that things *really* have these strange colours, which means that if you “and some group of distinguished researchers from your university“ would go there you would all agree that the sky is yellow, the grass is red and so forth. Furthermore, the intentional contents of attitudes and experiences regarding colour are also inverted. When a resident of Inverted Earth wonders, “why the sky is blue“ he is wondering why the sky is yellow and not why the sky is blue. One day a team of mad scientists knock you out, equip you with a set of colour inverting lenses and change your body pigments so that your body will have its usual colour when you look at it through your new lenses. Then they transport you to Inverted Earth where you are substituted for someone who has occupied a niche on Inverted Earth that corresponds exactly to your place back home (with the exception for the colour of things). When you wake up you will not know that you have been moved. What will happen? As far as the qualitative aspect of your mental life is concerned, nothing is different, Block says. E.g., you will still have “blue-experiences” when you look at the sky. What about the representational content of your perceptions? Block foresees the following: the first day on Inverted Earth, the intentional content of your colour experience remains the same, that is, different from the others. When you think, “the sky is certainly blue today”, you will be wrong. You will be applying the Earth term “blue“, which means blue, to the yellow sky. But after a period of time, you will be embedded in the linguistic environment of Inverted Earth and you will adopt the same intentional content as the rest of the inhabitants on Inverted Earth, and then you will be right when you think that the sky is blue. But the phenomenal character of your experiences will not change! You will still have “blue-experiences” whenever you look at yellow things, Block argues.<sup>82</sup>

How can a phenomenal externalist respond to this? She can deny that the properties your experiences represent on Inverted Earth (the yellow ones) will appear blue to you. Or she can deny that your yellow-caused

---

<sup>82</sup> The fact that you at some point will change from being wrong to being right opens the way to objections. In order to avoid these objections Block later changed the intra-personal example for an inter-personal one. Instead we are to imagine a pair of genetically identical twins. One of them is adopted on Inverted Earth after having inverted lenses inserted and the other grows up on Earth without any special lenses. The twins have the same qualitative contents at every moment, but are functional and intentional inverses of each other when it comes to colour experiences.

experiences will represent yellow on Inverted Earth. Michael Tye goes for the second option. He denies that the earthling's yellow-caused experiences on Inverted Earth represent yellow, no matter how long he stays there. Inverted Earth is not this person's "natural habitat", and therefore the conditions there are not optimal, and for this reason Tye thinks he can deny that the colour her brain state *tracks* on Inverted Earth is the colour it *represents*.

"Anyone who thinks that representational content is ever determined (fully or partly) by causal covariation or tracking needs to index the covariation or tracking to some particular range of conditions. What matters intuitively is the tracking that *would* obtain, *were* the conditions normal or optimal."<sup>83</sup>

Lycan seems to go for the other option. Even though my perceptions on Inverted Earth represent the real colours of things, I fail to be introspectively aware of this. Lycan acknowledges that if he was transported to Inverted Earth he would not notice a change from an introspective point of view, even though the representational content of his perceptual states has changed.

"The same goes for propositional attitudes, i.e., the Earthling transported to Twin Earth would notice nothing introspectively, despite the change in her/his belief and desire contents. Yet the attitude contents are still wide. Wideness does not entail introspective change under transportation."<sup>84</sup>

So while the perceptual experiences of the person who has been moved to Inverted Earth represent the property yellow, the person fails to be introspectively aware of this fact.

Now, while the Twin Earth example concerns the meaning of people's utterings, the Inverted Earth example primarily concerns the content of people's perceptions. Block putatively showed that the phenomenal character of perceptual experiences is not entirely determined by whatever colours the external objects may have. The truth value of reports about the phenomenal aspects of perceptions is, if Block is right, not determined by how things stand in the world, and first person authority is hence not

---

<sup>83</sup> Tye 1998, pp. 682-683.

<sup>84</sup> Lycan 2001c, p. 24.

threatened in this sense. When I, who live on Inverted Earth, say, “The sky seems blue to me”, I will be correct since my perception of the blue sky has the intrinsic property “bluish” (or something like that) to which I am referring. The phenomenal externalist seems to be forced to claim that a subject does not enjoy first person authority concerning the character of her perceptual experience. If the environment suddenly has changed without the subject’s knowledge, so have her experiences since they are determined by external circumstances.

## 2.4 “FIRST PERSON REPORTS” ARE NOT INTROSPECTIVE REPORTS

Armstrong proposes that sentences like “I am in pain now”, are not always used in a cognitive sense. It can even happen that they do not express any beliefs at all.<sup>85</sup> The sentence “I am in pain” can be used as a substitute for a groan or a cry for help. If reports about people’s sensations are used in this non-cognitive way, it makes little sense to speak of them as being either right or wrong. This fact may be misleadingly put as: it is logically impossible to be mistaken about such utterances. He writes:

“So, when philosophers have considered sentences like ‘I am in pain now,’ they have been misled by their ambiguity. They have moved between their noncognitive use, where the question of intellectual mistake does not come up but equally there is not question of cognition, and their autobiographical use, where there is no doubt cognitive certainty but simply *empirical* certainty. And so we persuade ourselves that such sentences express reports, but reports that have a special certainty and special authority.”<sup>86</sup>

Armstrong also claims that a sentence like “It looks green to me” may not be about any of the speaker’s *experiences*, it could express a tentative belief that something in the environment is green. This point is probably adopted from Wilfred Sellars who argues that talk about appearances is parasitic on talk about how the world is.<sup>87</sup> According to Sellars, the function of using

---

<sup>85</sup> Armstrong, 1963.

<sup>86</sup> Ibid., p. 429.

<sup>87</sup> Sellars 1956.

sentences about how things “look” is to withhold the commitment to the world that comes with talking about how things “are”. The mythical tie salesman Jones discovers that his colour judgement becomes fallible in interior lightning and therefore shifts from saying, “it is green” to “it looks green” to his customers. So when we say that the tie looks green, we are still talking about the tie, and maybe how *it* appears to us, but not about our experiences of the tie.

## 2.5 SUMMARY OF CHAPTER TWO

The infallibility thesis originates from the assumption that an experience is a private thing to which only the subject can have access. One typical objection to this is that even if experiences were private in this sense, it would not follow that the subject’s beliefs about these private experiences are infallible, it only follows that her beliefs are incorrigible, since no one else can inspect her experiences. The infallibility thesis, when it was “the official doctrine “as Ryle called it in *The Concept of Mind*, was most of the time accompanied by the notion that self-knowledge is *immediate*. But the assumption that self-knowledge is immediate is also sometimes advocated by philosophers who reject both the infallibility and the incorrigibility thesis of first person knowledge. The explanation might be that if we reject incorrigibility as well as infallibility there is really nothing that distinguishes introspective beliefs from other kinds of (empirical) beliefs. *If* beliefs about occurrent mental states are corrigible, is it then a fact that these beliefs do not differ in epistemic status from beliefs to the effect that similar states exist in other people? Audi claims that a number of contrasts remain, one of them being that first person knowledge is immediate while knowledge of other people’s mental states is not. And, he concludes: “other things being equal, direct beliefs are more reliable than indirect.”<sup>88</sup> As we have seen there are philosophers who argue that first person knowledge is fallible, corrigible as well as mediated. Lycan is one of them.

The problem of justification of introspective beliefs has also been touched upon. Some philosophers have maintained that such justification is impossible, since a person’s knowledge about her own experiences is not

---

<sup>88</sup> Audi, 1993, p. 179.

based on anything else than the presence of the experience itself. But from this fact different conclusions have been drawn:

(i) *S can never be wrong about P implies S has infallible knowledge about P.*

Or:

(ii) *S can never be wrong about P implies S does not know anything about P.*

Recently, a metaphysical theory that claims that beliefs and perhaps experiences are constituted by relational facts about a person has become influential. The doctrine of first person authority and that of semantic externalism have proven difficult to combine. Dretske's solution to this putative problem has been to limit first person authority to knowledge about what you think or perceive. Both he and Burge argue that we know as little about our introspective beliefs as we do of our perceptual beliefs.

### 3. The introspective process

The problem of the epistemic status of introspective beliefs is obviously related to the question of what it is we introspect – the object of introspection. Both issues are related to whatever conceptions we have of the introspective process. If one e.g., thinks that the object of introspection is inner in a metaphysical sense, one also tends to think that introspective knowledge is immediate and infallible (a feature that partly defines the inner world). The introspective process may then be regarded as a kind of inner perception, where the subject has unmediated access to her own experiences. If one, on the other hand, thinks that the object of introspection is inner in a mere spatial sense - it is inside the body - one might also picture introspective knowledge to be like perceptual knowledge, but in a different sense. It is fallible just like empirical knowledge in general. The introspective process can still be called “inner perception” but it is then regarded as a causally mediated process, just like external perception is. Introspection can thus be likened to perception in many different ways. Below we will briefly go through some of the most predominant understandings of introspection as a process. First a general note about perception.

#### 3.1 PERCEPTUAL AWARENESS

A familiar fact about ordinary perception is that it does not always produce true beliefs about the world. Perceptual beliefs are sometimes false. This is a feature which Lycan and Armstrong take introspection to exhibit as well, while Brentano e.g., disagrees and claims that inner perception is always accurate. However, opinions differ among philosophers about whether the result of a perceptual process always is a belief. Most philosophers agree that perception can either make us aware that a certain fact obtains, or just make us aware of some object and/or the properties of that object. I can be aware of a tingling in my foot or aware of the sound of a fan, while not necessarily believing that there is a tingling in the foot or that there is a fan that makes a sound. Some philosophers, however, like Daniel Dennett think of perception entirely in cognitive terms. There is according to him nothing like “object-

awareness”, perception is always “fact-awareness”. Perceptual experiences are so to speak “mini”- or “proto-beliefs”.

In discussions of the epistemic status of introspection, it is implied that introspection gives rise to beliefs which can be true or false. But is it not also possible that introspection is like perception in this sense and that introspection may sometimes involve mere object-awareness? It may be fruitful to have this distinction in mind in the following discussion.

### 3.2 FOUR KINDS OF INNER PERCEPTION

Many philosophers have talked about “inner perception” and an “inner sense”. These terms have, however, been used with different meanings by different philosophers.<sup>89</sup> The following main understandings can be distinguished:

- (i) Inner perception as a special kind of obligatory awareness of one’s own mental states.
- (ii) Inner perception as a deliberate act in which the subject actively reflects on or attends to some of her mental processes.

Wilhelm Wundt labelled these processes, *innere Wahrnehmung* (inner perception), (i) and *Selbstbeobachtung*, which is introspection, (ii).

---

<sup>89</sup> It is argued by many philosophers that Kant is the originator of the inner sense theory. Kant’s theory is however much too complex to go into here. Andrew Brook (2001) says that Kant’s theory of inner sense is “a mess”. (p 16). He continues:

“Here are just a few of the problems. Kant insists that all representational states are in inner sense, including those representing the objects of outer sense (i.e., spatially located objects), but he also says that the object of inner sense is the soul, the object of outer sense the body (including one’s own). He comes close to denying that we can be aware of the denizens of inner sense – they do not represent inner objects and have no manifold of their own. Yet he also says that we can be aware of them – representations can themselves be objects of representations – and that representations can make us aware of themselves. In its role as a form of our means to awareness of self, apperception ought to be part of inner sense. Yet Kant regularly contrasted apperception, a means to awareness of oneself and one’s acts of thinking, with inner sense as a means to awareness of – what?” (pp.16-17.)

### 3.2.1 Awareness of awareness

Is there a difference between being aware of, say a sound and being aware of oneself hearing the sound? Semantically, there is, but are there also two different processes at work here? Is it possible to be aware of *just* the sound and not of *the hearing of the sound*? Can we also be aware only of the hearing, but not of the sound? Franz Brentano claims that the distinction between hearing a sound and being aware of hearing the sound is legitimate, even though the two acts always appear together *in reality*. Some contemporary philosophers, Armstrong and Lycan being two of those, agree with Brentano that there is an important and real difference between being aware of a physical object (like a sound) and being aware of the perception of the object (like the hearing of the sound) but they think, in contrast to Brentano, that awareness of the perception is not always present but only accompanies our experiences occasionally.

Lycan and Armstrong on the one hand, and Brentano on the other, have different views about how this self-awareness is constituted. Armstrong and Lycan argue that self-awareness is constituted by one mental state being about another. Brentano dismisses such a view since he thinks it implies an infinite number of acts being conscious of other acts. Brentano presupposes that a mental act must be a conscious act, and if “conscious act” means “act the subject is conscious of having”, we do not only need a second order conscious act to be conscious of the first order act, but also a third order conscious act to be conscious of the second order act, and so on. Lycan and Armstrong argue that the second order state, which makes the subject aware of the first state *need not itself be a conscious state*. Brentano is however reluctant to allow for the possibility of unconscious mental states. He presents several arguments against postulating “unconscious conscious mental states”.<sup>90</sup> Instead he tries to solve this problem by claiming that one and the same act has two different intentional objects. Every intentional experience has a “double object”, a primary and a secondary one.

“The presentation of the sound and the presentation of the presentation of the sound form a single mental phenomenon; it is only by considering it in its relation to two different objects, one of which is a physical phenomenon and the other a mental phenomenon, that we divide it conceptually into two presentations. [...] We can say that the sound is the

---

<sup>90</sup> See Brentano 1995 [1874], chapter 2.

*primary object* of the *act* of hearing and that the act of hearing itself is the *secondary object*.”<sup>91</sup>

Since Armstrong and Lycan are operating in a time when the existence of unconscious mental states are widely accepted, they do not feel the need to refrain from proposing the existence of higher order mental states in order to explain what makes a conscious state conscious.

Brentano is by some philosophers<sup>92</sup> pictured as the originator of the so-called “higher order theory of consciousness”, defended by e.g., Armstrong and Lycan. There are, however important differences between Brentano’s theory and that of Armstrong and Lycan, as we have seen.<sup>93</sup> It is true that Brentano identifies “the mental” with intentionality or “consciousness of”. He writes for example: “We have seen that no mental phenomena exists which is not, in the sense indicated above, consciousness of an object...”<sup>94</sup> But at the same time he is not ready to accept a theory that postulates *acts being conscious of other acts*. Lycan emphasises the independence of states of the first and the second order and allows for the existence of unconscious states, states that are not, in turn, the objects of a higher order state. Brentano, on the other hand, maintains that the awareness that makes an act conscious is a dependent aspect of the original mental act and not a new act. Temporally, both acts occur at the same time, but ontologically, the primary awareness of the sound is prior. This leads Brentano to allow for the possibility that a mental state could exist without this dependent aspect but that the reverse is inconceivable.

“A presentation of the sound without a presentation of the act of hearing would not be inconceivable, at least *a priori*, but a presentation of the act of hearing without a presentation of the sound would be an obvious contradiction.”<sup>95</sup>

Husserl denies that there is any phenomenological evidence in support of there being an obligatory and continuous inner perception. He claims that mental acts are simply lived through as moments of experiences; they are as a rule not experienced as intentional objects of another act (or the same act).

---

<sup>91</sup> Brentano 1995 [1874], pp. 127-128.

<sup>92</sup> Güzeldere 1997, p. 789, Siewert 1998, p. 357 n1 and 358 n3.

<sup>93</sup> This is also pointed out by e.g., Thomasson 2000, and Zahavi 2004.

<sup>94</sup> Brentano 1995 [1874], p. 102.

<sup>95</sup> *Ibid.*, p. 128.

“Inner perception” only happens in psychological reflection, he says. We can direct our attention towards our experiences, and then these experiences will become the intentional objects of an inner perception, but this only occurs when we actively reflect upon them. Husserl distinguishes between *perceiving* (Wahrnehmen) and *experiencing* (Erleben). Prior to reflection one perceives a perceptual object, but one does not perceive the perception, one experiences the perception. Even when I am not directed towards the perception in any sense, the perception is conscious in the sense of being experienced.

Brentano argues that although we constantly perceive our mental processes, we cannot *actively observe* them, we cannot reflect upon them, (in so far as reflection implies attentive perceiving, which was taken for granted by Brentano as well as many others of his time). Introspection, taken as “attentive perceiving” is hence not possible according to Brentano.

“Psychology, like the natural sciences, has its basis in perception and experience. Above all, however, its source is to be found in the *inner perception* of our own mental phenomena. We would never know what a thought is, or a judgement, pleasure or pain, desires or aversions, hopes or fears, courage or despair, decisions and voluntary intentions if we did not learn what they are through inner perception of our own phenomena. Note, however, that we said that inner *perception* [*Wahrnehmung*] and not introspection, i.e. inner *observation* [*Beobachtung*], constitutes this primary and essential source of psychology. These two concepts must be distinguished from one another. One of the characteristics of inner perception is that it can never become inner observation. We can observe objects which, as they say, are perceived externally. In observation, we direct our full attention to a phenomenon in order to apprehend it accurately.”<sup>96</sup>

The terms used by Brentano and Husserl respectively to account for these different processes may lead to some extra confusion. Brentano calls the pre-reflective obligatory self-awareness “inner perception” or “*Wahrnehmung*”, while Husserl uses the same term for the attentive, deliberate kind of self-awareness, which, in turn, is called “*Beobachtung*” by Brentano. Husserl’s “*Erleben*” is a third concept. To “*erleb*” some mental act or state is simply to undergo it (consciously), but it does not entail that the subject is conscious *of* the act in the sense Brentano took it to be.

Lycan, again, thinks that both passive inner perception as well as active, attentive inner perceiving (by him called “introspection”) is possible.

---

<sup>96</sup> Ibid., p. 29.

According to him the same mental state can be either unconscious or conscious. In the latter case it is represented by a higher order mental state. A special case is when it is *introspected* which means that the subject not only perceives it, but also attends to it (perceptually). So while Brentano argues that every mental act is constantly and passively innerly perceived (by itself) and that attentive self-perception is impossible, Lycan maintains that only some mental states are innerly perceived by other mental states, and that some of this inner perceiving is attentive and hence comprises introspection. Armstrong makes the same distinction as Lycan does, but speaks of “reflex introspective awareness” and “introspection proper”.<sup>97</sup>

John Locke has a reputation for having founded the theory of inner sense. He claims that ideas are the materials of knowledge and that all ideas come from experience. Experience is of two kinds, sensation and reflection. One of these - sensation - tells us about things and processes in the external world. The other - reflection - tells us about the operations of our own minds. Reflection is a sort of internal sense that makes us conscious of the mental processes we are engaged in. About reflection he writes:

“Secondly, the other fountain, from which experience furnisheth the understanding with ideas, is the perceptions of our minds within us, as it is employed about the ideas it has got; which operations, when the soul comes to reflect on and consider, do furnish the understanding with another set of ideas which could not be had from things without: and such are perception, thinking, doubting, believing, reasoning, knowing, willing and all the different acting of our own minds; which we being conscious of, and observing in ourselves, do from these receive into our understanding as distinct ideas, as we do from bodies affecting our senses. This source of ideas every man has wholly in himself: and though it not be sense, as having nothing to do with external objects, yet it is very like it, and might properly enough be called internal sense. But as I call the other Sensation, so I call this REFLECTION, the ideas it affords being such only as the mind gets it by reflecting on its own operations within itself.”<sup>98</sup>

It seems probable that Locke only recognized the reflective, active kind of inner perception, that which Brentano calls “inner observation”, and Lycan and Armstrong “introspection (proper)”.

---

<sup>97</sup> Armstrong 1980, p. 63.

<sup>98</sup> Locke 1924 [1690], pp. 43-44.

### 3.2.2 Direct and indirect inner perception

Now, these views about inner perception also differ in their epistemic considerations. According to Brentano inner perception is, unlike outer perception, infallible and he claims that it is inner perception alone that possesses this characteristic. Brentano ultimately draws the conclusion that inner perception is the only kind of perception which is worthy of its name. He writes:

“[I]nner perception is not merely the only kind of perception which is immediately evident; it is really the only perception in the strict sense of the word. As we have seen, the phenomena of the so-called external perception cannot be proved true and real even by means of indirect demonstration. [...] Therefore, strictly speaking, so-called external perception is not perception. Mental phenomena, therefore, may be described as the only phenomena of which perception in the strict sense of the word is possible.”<sup>99</sup>

And Husserl writes (about inner attentive perception, the kind he recognises):

“The ‘self-evidence’ usually attributed to inner perception, shows it to be taken to be an *adequate* perception, one ascribing nothing to its objects that is not intuitively presented, and given as a real part (*reell*) of the perceptual experience. [...] It is accordingly clear, and evident from the mere essence of perception, that adequate perception can only be ‘inner’ perception, that it can only be trained upon experiences simultaneously given, and belonging to a single experience within itself.”<sup>100</sup>

According to Armstrong-Lycan inner perception is just like outer perception and hence fallible. We may be mistaken about the states we are introspectively aware of. Armstrong writes:

“Eccentric cases apart, perception, considered as a mental event, is the acquiring of information, or misinformation about our environment. It is not an ‘acquaintance’ with objects, or a ‘searchlight’ that makes contact with them, but is simply the getting of beliefs. Exactly the same must be said of introspection.”<sup>101</sup>

---

<sup>99</sup> Brentano 1995 [1874], p. 91.

<sup>100</sup> Husserl 2001 [1900/1901], p. 86.

<sup>101</sup> Armstrong 1968, p. 326.

As we can see, while Brentano thinks that inner perception is unlike outer perception in that it is infallible and direct, Armstrong and Lycan argue that it is like outer perception, i.e., fallible and mediated, only “inner”.

We can now extract four notions of inner perception: (i) passive inner perception in Brentano’s version, (ii) active inner perception in Husserl’s version, (iii) passive inner perception in Lycan’s and Armstrong’s version and (iv) active observation in Lycan’s and Armstrong’s version. (iii) and (iv), which we will have reason to come back to, will from now on be called: *NAHOP* for “non-attentive higher order perception”, and *AHOP* for “attentive higher order perception”.

### 3.3 INNER THINKING

An alternative to the higher order *perceptual* model of Lycan and Armstrong’s has been put forward by Rosenthal. He claims, just as Lycan and Armstrong do, that a conscious state is a state which is the intentional object of another mental state. However, he denies that the process which brings about this consciousness is similar to perception. He argues that the subject becomes conscious of her mental states by *thinking* or believing that she is in those states, not by perceiving them.

While Lycan argues that active introspective observation is an attentive kind of second-order inner perceiving, Rosenthal claims that introspective beliefs are beliefs that are about second order beliefs.

Peter Carruthers puts forward another version of the higher order thought theory. This theory claims that in order for a mental state of the first order to be phenomenally conscious, it only needs to be *available* for a higher order thought. It is, hence, sufficient that a first order experience is disposed to cause a higher-order belief that the subject is having this experience, for the subject to be conscious of this experience.<sup>102</sup>

---

<sup>102</sup> Carruthers 2000a, 2004. Carruthers claims that his theory also can count as a kind of higher order perception theory.

### 3.4 DISPLACED PERCEPTION

Dretske presents an account of the introspective process, which also makes use of the term “perception”. He claims that introspection is in some important aspects similar to what he calls “displaced perception”. To perceive something in this displaced manner is to become aware that something is the case, not by seeing it directly, but by seeing something else. A example of this would be that one can see that the gas tank is empty by looking at the gas meter. The point is that we become aware of some fact pertaining to an object, not by being *aware of* that object, but by directly perceiving another object. Dretske argues that introspective beliefs are acquired in a similar way. We become aware of what our experiences are like, not by attending to them, and not even by “looking at them from the corner of our eye”, but by observing something else, namely that which they are experiences of, i.e., physical objects, their properties and facts about them. We can thus, according to Dretske, become aware of what our perceptual states are like, by attending to what they are perceptions of. Tye advocates a similar theory and so does Gareth Evans.<sup>103</sup>

### 3.5 REPLAY OF PERCEPTION

William Lyons claims that introspection should be described as a *replay of perception*.

“[W]hen we ‘introspect’ in order to discover things about our cognitive, appetitive, and affective lives, we engage in a process of perceptual ‘replay’. We ‘replay’ or recreate – at least with ‘edited highlights’ or in ‘dramatized form’ – what we think we said, or would have liked to have said, or did or felt.”<sup>104</sup>

This account has, in fact, little to do with any of the theories of inner perception previously accounted for. Lyons emphasises that his account of introspection does not in any way assume that introspection is a “meta-process” that monitors other mental processes, but that it is these other processes put *to a certain use*.

---

<sup>103</sup> Tye (2000b), Evans (1982).

<sup>104</sup> Lyons 1986, p. 113.

“‘[I]ntrospection’ is our employment of perceptual memory and imagination to find out about our motives, thoughts, hopes, desires, and the like by finding out about our published, revealed-in-speech-gesture-expression-and-behavior motives, thoughts, hopes, desires, and so on, for these are all we have conscious access to.”<sup>105</sup>

Instead of monitoring what is going on at “the first level of consciousness” as Armstrong and Lycan assume, introspection “substitutes a working model or dynamic picture for what it cannot know first hand.”

“We might think that we are catching ourselves monitoring, inspecting, scanning our acts of thinking, hoping, deciding, loving, or needing when in fact all that we are doing is replaying internally to ourselves or imaginatively constructing for ourselves a model version made out of overt behavior, gesture, and expression that we describe in our ‘folk psychology’ as thinking, hoping, deciding, loving and needing.”<sup>106</sup>

The object of the introspective process in this case is not an experience, in fact, there is no proper object at all. The introspective process does not capture any mental objects, nor is it caused by mental objects. It “invents” a mental life as it goes along.

Lyons’ theory of “introspection” does not accord with my preliminary characterisation, in so far as the object of introspection according to him are not episodic experiences. His main motive for evoking this kind of theory is to avoid a “two level account” of introspection, where one level or part of consciousness attends to or perceives another. There are several excellent reasons why a two level theory should be rejected, even though Lyons himself does not mention them (he brings forward other less successful counter-arguments though).

---

<sup>105</sup> Ibid., p. 114.

<sup>106</sup> Ibid., p. 125.

### 3.6 SUMMARY

The introspective process is often modelled as a kind of inner perception, but as we have seen, “inner perception” does not always denote the same thing. The difference between Lycan’s and Brentano’s account lies both in the issue of allowing for different orders of mental states and in the epistemological standpoint. Brentano claims that the act is aware of itself, while Lycan argues that a conscious state is a state that is the object of yet another mental state. While Brentano argues that inner perception is infallible, Lycan thinks it is fallible, corrigible, and mediated both in a metaphysical and epistemological sense. Displaced perception may be regarded as even more indirect than the introspective process according to Armstrong-Lycan, since we are not at all *aware of* our experiences, but only aware that we have them via being aware of something else. Lyons, who also makes use of the term “perception” has yet something else in mind.

### 3.7 FINAL COMMENTS

The objective of the first part of this thesis, has been to give a general account of what different philosophers have in mind when they talk about introspection. It has been my intention to briefly summarise some influential views of introspection, not to take a stand for or against any specific account. I have however not been able to stop myself from some occasional remarks.

The survey has been divided into three different parts; the object of introspection, the epistemology of introspective beliefs and the nature of the introspective process. This partition does however not entail that these aspects can be fully understood in separation from one another.

In the discussion about the object of perception we have distinguished between a dualist view of what “the inner” is and a materialist view. A dualist theory of mind has no problems allowing for the existence of intrinsic sensory qualities of experience, intrinsic sensory qualities that are more or less by definition accessible by introspection. Therefore the need for determining which kind of process makes us aware of these objects may not be felt as pressing. A materialist version of a theory of mind must possibly reject the existence of such special introspectible mental qualities and may instead claim that sensory qualities are really part of the intentional content

of an experience. If one argues that it is the intentional contents that are the objects of introspection, and especially if this claim is combined with an overall materialist theory of mind, the question of what (if anything) makes the subject's awareness of these objects special becomes more urgent.

It may seem more natural to assume that introspective knowledge is fallible if one holds on to a materialist theory of mind than if one is a dualist. In the first case there are no distinguishing marks that automatically make introspective knowledge epistemically different from ordinary perceptual knowledge. Cartesian dualism, on the other hand, is partly defined in epistemic terms. The mind is that of which the subject has the most secure knowledge. So while the idea of object, epistemology and process are at least partly independent on one another, there are some combinations of views which may seem more intuitively appealing: the view that the mind is material in combination with the claim that knowledge of the mind is fallible and perception-like in an indirect and mediated sense, or the idea that the mind is immaterial, in combination with the claim that we have infallible knowledge of it, and that the knowledge process is perception-like in an direct and unmediated sense. But philosophers have also maintained that while the mind is material, the subject's knowledge of her own mind is infallible.



Part two:  
representation,  
meta-representation  
and introspection

The focus of the remaining part of this thesis will be on a specific group of contemporary theories of introspection. These are the theories of David Armstrong, Peter Carruthers, William Lycan, David Rosenthal, Fred Dretske, Sydney Shoemaker and Michael Tye. The theories can be divided into two main groups, one which is usually referred to as that of “higher order theories” (HO) and one which we will call “one level theories” (OL). The HO theories entail that there are mental states of different orders, there are first order states (e.g., perceptions) that are about the world and second order states that are about the first order states while the OL theorists claim that there is only one level of states, namely states that represent the world. Introspective states should according to the latter philosophers not be described as being of a higher order since these states do not make the subject *aware of* her mental states; the result of an introspection is *awareness that* one is in a certain state. The first four philosophers mentioned above are HO theorists, the latter three are proponents of versions of the OL theory.

One of the main objectives of all these theories (it may even be *the* main objective) is to give a naturalistic account of phenomenal consciousness. This is done by appealing to intentionality and to some extent, to functionality. The higher order theories (HO) do not only claim that introspection is some kind of inner awareness of first order states, but also that phenomenal consciousness is explained by such awareness. It is hence not possible to analyse the different accounts of introspection and at the same time avoid discussing the claims about phenomenal consciousness made by the HO theorists.

The main focus of these theories when it comes to introspection is introspection of perceptual experiences. Possibly this is so because the thesis that our thoughts and perceptions are entirely explainable in representational (and functional) terms faces its biggest challenge when it comes to accounting for perceptual experiences. As was discussed in the first part of the thesis, one of the most controversial and most problematic claims made in these discussions is that of phenomenal externalism. The theories that will concern me the most below hold that the qualitative character of perceptual experiences is determined by their wide contents.

The HO and the OL theorists have been involved in a continuous discussion with one another, both regarding the nature of introspection and about how to explain phenomenal consciousness. In this part I will first briefly describe an underlying assumption of most of these theories – naturalistic representationalism about mental states. After that, the different

accounts of introspective awareness will be briefly presented and the differences in the accounts of phenomenal consciousness will be described. Then, I will consider some arguments for and against them and attempt to show how and why neither of the theories is able to give successful accounts of introspection.

## 4. Representational theories of consciousness and introspection

The so-called *representational theories of consciousness* have gained increasing popularity during the last couple of decades. The origin of these theories can however be traced back to Brentano. Brentano claims that the mark of the mental is intentionality. All mental acts are about something. (They are, according to him, partly about themselves and partly about something else.) The modern version of this theory assumes that intentionality is representation. A mental state represents some object, fictitious or real, and sometimes a state of affair that either obtains or does not obtain. Unlike Brentano, the modern representationalists often claim that the *phenomenal character* of being in a particular state can also be understood as part of the representational content of the experience. The perception of a red cow is a representation of a cow that is red. A subject's experience of red is accounted for in terms of her being in a perceptual state that represents a certain property of an external object. In the same way as e.g., a red-experience is an experience that represents something red, a pain-experience should be described as an experience that represents tissue damage, or C-fibre stimulation (or something like that); an experience of sour represents the sourness of some physical object, and so on. According to some philosophers, the functional role a certain experience plays also contributes to "what it is like" to have this experience. Dretske, Tye and Lycan describe this thesis in the following ways:

"If, in accordance with the Representational Thesis, we think of all mental facts as representational facts, the quality of experience, how things seem to us at the sensory level, is constituted by the properties things are represented as having."<sup>107</sup>

"Phenomenal character (or what it is like) is one and the same as a certain sort of intentional content."<sup>108</sup>

---

<sup>107</sup> Dretske 1995, p. 1.

<sup>108</sup> Tye 1995a, p. 137.

“[T]he mind has no special properties that are not exhausted by its representational properties, along with or in combination with the functional organization of its components.”<sup>109</sup>

This kind of theory will in the following will be referred to as “representationalism”. The main alternative to representationalism is in contemporary discussions usually called “phenomenalism” or “phenomenism”. Phenomenalism rejects the representational thesis and maintains that the phenomenal character of experiencing is not exhausted by the representational content of perceptual states together with their functional role.

The representationalist thesis is accepted in different degrees by different thinkers. Some representationalists argue that differences in experiential character can be accounted for in representationalist terms *only within one given sensory modality*. So the overall difference between e.g., tasting a tomato and seeing it cannot be reduced to a difference in intentional content. The felt difference needs to be accounted for in some other way, preferable in terms of functionality. The visual representation of the tomato and the gustatory representation play different functional roles in the subject’s mind and this is experienced from a subjective point of view as different *modes of presentation*. A stronger version of representationalism claims that every qualitative difference, even between experiences in different sense modalities, is a difference in intentional content. The felt difference between tasting a tomato and watching one is entirely a matter of what the two experiences represent.<sup>110</sup> Lycan is a proponent of the former version, he appeals to functional facts in order to explain inter-modal qualitative differences. Shoemaker recommends a similar view. Dretske and Tye, on the other hand, are advocates of strong representationalism. They argue that all differences in phenomenal character are representational differences.

A representationalist theory of consciousness need not be a reductive theory. One can plausibly argue that qualia are identical with intentional content, and at the same time maintain that qualia demand an ontologically unique kind of representation, a kind that cannot be reduced to any physical or functional properties. The theories that will concern us here are however

---

<sup>109</sup> Lycan 1996, p. 11.

<sup>110</sup> The distinction between strong representationalism and the version which also takes functional facts into account is by Block (1996, pp. 37-38) described as one between *quasi-representationism* and *representationism* and by Byrne (2001b, p. 205) as one between *intra-modal* and *inter-modal intentionalism*.

all naturalistic theories. (Part of) their overall aim is to explain phenomenal consciousness in naturalistic terms.<sup>111</sup>

There is also a distinction to be made between phenomenal *externalism* and phenomenal *internalism*. The former thesis entails that the phenomenal character of experiences is entirely a matter of the experience representing objective properties of physical objects. Internal representationalism on the other hand, holds that the phenomenal character of experience consists in its “narrow content”, it is “in the head”, in Putnam’s words. Hence for the externalist, but not for the internalist, the truth conditions of statements like “I seem to see something blue” involve factors external to the perceiver. Dretske, Lycan and Tye are phenomenal externalists. Shoemaker and Carruthers defend the internalist version. They claim that the properties represented are phenomenal. An externalist may agree with them that a physical object like a blueberry is both objectively and phenomenally blue, but still holds that the property represented is the objective one.

The main focus here will be on representationalist theories. There are however some other versions of the HO theory that will be considered here. Armstrong’s theory of introspection is relevant to this thesis since he also claims that consciousness can be explained by higher order perceptions of our own mental states. He is however not a representationalist, but argues that mental states should be analysed in terms of their causal relations to behaviour. Rosenthal’s theories of introspection and phenomenal consciousness will also be discussed here since he is a strong proponent of a version of the HO theory, but he does not at all think that phenomenal character is reducible to the intentional content of perceptual experiences.

#### 4.1 PHENOMENAL CONSCIOUSNESS

One of the major claims of the representational theory of mind is that the problem of *phenomenal consciousness* can be reduced to the problem of representations. What is the problem of phenomenal consciousness? Philosophers who in different ways try to naturalise the mind tend to think that this aspect is the hardest to account for. As David Chalmers puts it:

---

<sup>111</sup> Lycan says on many occasions that he does not propose to *explain* phenomenal character, sensory qualities, or qualia. See e.g., Lycan 2004, p 96. But what he does try to do is to give an account of phenomenal consciousness in terms of awareness of those qualities.

“Even when we have explained the performance of all the cognitive and behavioural functions in the vicinity of experience – perceptual discrimination, categorization, internal access, verbal report – there may still remain a further unanswered question: Why is the performance of these functions accompanied by experience? [...] This further question is the key question in the problem of consciousness.”<sup>112</sup>

And Rosenthal writes:

“Many would accept the possibility of an essentially nonmental explanation of what it is for states to have intentional properties, at least in the case of nonconscious intentional states. And [...] perhaps the same is true of sensory states as well. But it is far less likely that we can successfully do this in any direct way with consciousness. Nothing in nonmental reality seems suited to explain what it is for a mental state to be conscious”<sup>113</sup>

These two quotes reflect the consideration that neither functionality nor intentionality can explain phenomenal consciousness. So what is “phenomenal consciousness”? Ned Block introduced the term in the modern discussion. He picks out two meanings of consciousness, *phenomenal consciousness* and *access consciousness*, and describes them as follows:

“P-consciousness [phenomenal consciousness] is experience. P-conscious properties are experiential properties. P-conscious states are experiential states, that is, a state is P-conscious just in case it has experiential properties. The totality of the experiential properties of a state are ‘what it is like’ to have it.”<sup>114</sup>

“A is access-consciousness. A state is A-conscious if it is poised for direct control of thought and action. To add more detail, a representation is A-conscious if it is poised for free use in reasoning and for direct ‘rational’ control of action and speech.”<sup>115</sup>

---

<sup>112</sup> Chalmers 1998, pp. 4-5.

<sup>113</sup> Rosenthal 1997, p. 735.

<sup>114</sup> Block 1997, p. 380.

<sup>115</sup> Ibid., p. 382.

The general idea is that while A-consciousness can be accounted for in physical and/or functional terms, there does not seem to be a neat way to do the same with P-consciousness.

The meaning of “phenomenal consciousness” and its cognates cannot easily be captured. Different philosophers favour different descriptions of this special way of being conscious, as e.g., “what it’s like”, “the subjective dimension”, “the first person perspective”, “experience” and so on. The HO theorists claim that phenomenal consciousness should be interpreted in terms of subjectivity. What it is like to consciously perceive a red rose should be understood as “what it is like *for me* to perceive a red rose.” Lycan describes “subjective consciousness” thus:

“A subject’s consciousness in this sense is ‘what it is like’ for the subject to be in whatever mental state it is in. A tighter characterization might be: what can be described, if at all, only in the first person.”<sup>116</sup>

In the first quote from Block we find “phenomenality” ascribed both to consciousness as such and to properties of conscious experiences. The HO theorists argue that all perceptual states *have* experiential properties, but that the subject is not always phenomenally or subjectively conscious when she is in such a state. They claim that perceptual states are like something (in themselves) even when they are not like anything *for the subject*.

Phenomenal consciousness according to the HO theory consists in the subject *representing* or *being aware of* her perceptual states. Just being in a perceptual state enables the perceiving subject to have A-consciousness of the world. She will be both A-conscious and P-conscious when she is aware of her perceptual state. It needs to be added that the HO theorists do not use Block’s distinction in their discussions, but I think that this formulation fairly well captures the idea behind what they take the difference to be between those mental state which are phenomenally conscious those which are not.<sup>117</sup>

The OL theorists, on the other hand, claim that phenomenal consciousness should not be understood as a person’s being aware *of* her own experiences, but in terms of the intentional content of the perceptual

---

<sup>116</sup> Lycan 1996, p. 4.

<sup>117</sup> There is some (verbal) disagreement among the HO theorists. While e.g., Rosenthal call perceptual states that are not the intentional object of a proper higher order thought: “unconscious”, Carruthers claim that perceptual experiences exhibit “wordly subjectivity” and conscious perceptual experiences: “experiential subjectivity”.

experience and the special functional role these experiences play. They argue that phenomenally conscious experiences exhibit a certain sort of representational content (non-conceptual, systemic or analog content) and that it is the content as such that determines whether the subject is phenomenally conscious or not. States with this kind of content have a certain role in the functional system. In Tye's terms they are "poised to have an effect on the subject's behaviour". The presence of a certain kind of representational content in this sense is sufficient for phenomenal consciousness, the subject does not, in turn, have to be *aware of* these experiences.

## 4.2 QUALIA AS REPRESENTATIONAL PROPERTIES

What is a quale? There are important differences between a representationalist view on qualia, and a phenomenalist one. Below I will try to reconstruct the different views of what qualia are, taken by the philosophers under discussion.

Feelings and perceptual experiences seem to have some special qualitative character which they do not share with e.g., thoughts - it is "like something" to undergo them, in Nagel's terms.<sup>118</sup> The term "qualia" has been used to denote monadic properties which somehow make up this phenomenal dimension of experiences. According to C.I. Lewis's<sup>119</sup> sense of qualia, a quale is an introspectible monadic qualitative property of a phenomenal individual. According to a representationalist theory of mind, qualia belong to the representational content of experiences. If I am aware of a moving green object (like a John Deere tractor) my perceptual experience is not itself green or moving. It is the tractor that is green and my perception of the tractor represents the tractor as being green. We can, thus, hence discern the following two (rough) characterisations of qualia:

(i) Qualia are *intrinsic* features of experiences, which may vary without any variation in the intentional content of the experience. This view is upheld by e.g., Thomas Nagel (1974), Ned Block (1990) and David Rosenthal.

---

<sup>118</sup> Nagel 1974.

<sup>119</sup> Lewis 1929. This may be the original definition of qualia.

(ii) Qualia are *representational* properties of experiences. This is the view of e.g., Carruthers, Dretske, Lycan, Tye, and Shoemaker.

The notion of qualia as representational properties is complicated. To begin with: do the advocates of this view mean that qualia are *representing* properties of experiences or that they are *represented* properties of objects perceived? Are qualia properties of the representational vehicles or are they properties of the representational content? Dretske claims that qualia are part of the representational content in the sense that they are represented properties.

“In accordance with the Representational Thesis, I continue to identify qualia with phenomenal properties – those properties that (according to the thesis) an object is sensuously represented [...] as having.”<sup>120</sup>

However, in a later discussion, he writes:

“What this means is that *if* we follow philosophical convention and take qualia to be properties of one’s experiences (and not the properties one experiences), then it is  $\mathfrak{P}$ , not *P*, that is the quale.”<sup>121</sup>

In this paper, Dretske argues that if we insist on locating qualia inside the head, it follows that the subject is not aware of them when she is having a perceptual experience. According to this schema, qualia are thus representing properties, and not represented ones. Tye maintains that qualia are represented properties.

“Consider first the case in which I am having a visual experience of a red, round surface, as when I am viewing a ripe tomato. If you tell me to attend to aspects or qualities of my experience, I find myself attending to qualities that are not qualities of my experience at all – qualities such as redness and roundness. These are qualities that belong to the object of my experience.”<sup>122</sup>

Lycan’s view of qualia is complicated. To me it seems as if he is sometimes talking of qualia as represented properties and sometimes as representing

---

<sup>120</sup> Dretske 1995, p. 73.

<sup>121</sup> Dretske 1999, p. 113. Here “ $\mathfrak{P}$ ” stands for property of an experience and “*P*” for an experienced property, that is a property of an external object.

<sup>122</sup> Tye forthcoming (2006), p x.

properties. This ambiguity can be discerned if we compare the following two quotes:

“The Representational theory affords a third alternative, by supposing that qualia are *intentional contents* of sensory states, properties of intentional objects, represented properties of representata.”<sup>123</sup>

“I believe that introspection is the operation of an internal attention mechanism that monitors experiences and produces second-order representations of their properties, including especially their feature-representational properties.”<sup>124</sup>

In the first quote above Lycan is clearly talking about qualia as represented properties. In the second quote, he seems to slide over to describing qualia as representing properties, properties of the representations. We will discuss this issue in more detail below; for now it is enough to point out that this ambiguity seems to exist. Rosenthal, again, claims that qualia are non-representational properties of the perceptual experiences; by formulating a certain HOT the subject becomes aware of those properties. Shoemaker maintains that qualia are representing properties of experiences. He writes:

“Qualia can be thought of as the vehicles of the representation of properties of perceived objects. As such, they can be said to represent such properties.”<sup>125</sup>

Thesis (ii) above (p. 78) can thus be interpreted as either:

a) Qualia are represented properties of objects.

Or:

b) Qualia are representing properties of experiences.

Qualia can thus be thought of as being non-representing properties of the representational vehicle (Rosenthal), or they can be thought of as representing elements of the representational vehicle, elements in virtue of which the representation represents what it does (Shoemaker) or, finally,

---

<sup>123</sup> Lycan 2001c, p 18.

<sup>124</sup> Lycan 1999a, p 125.

<sup>125</sup> Shoemaker 2002, p. 468.

qualia can be thought of as properties of the object represented. Tye, Dretske and Lycan basically use “qualia” in the third sense, but Lycan sometimes seems to use it in the second sense and Dretske uses it in this sense in his discussion from 1999.

There are further ways of characterising qualia, such as in terms of the subject’s access to them. Those who claim (i) seem also to hold that qualia are immediately accessible to the subject. Lycan and Tye also think that the subject has access to the qualia she perceives. In his later discussion, Dretske rejects this claim. Qualia are representing properties and the subject is not aware of those properties, not when she perceives something and not even when she introspects her perceptions.

#### 4.2.1 What is being represented?

According to some of these theories perceptual experiences represent *physical* properties of objects (or the body). My experience of a red rose represents a red rose. And my feeling pain in the foot represents some damage in my foot. One slight problem with this account seems to be that roses and their likes are not straightforwardly red. And pains can be felt even when it is difficult to find some relevant tissue damage. Redness does not seem to be a property of a physical rose existing in this world (or in any other world) and pain is hard to account for as a property of the body. A scientific account of physical objects finds no place for qualities like colours.<sup>126</sup> Colours are according to this account not properties of objects or their surfaces, but come into existence with the help of the joint forces of light and a special kind of sensory organ which some creatures have. These facts have led some philosophers to treat colours as *dispositional properties* of physical objects. A red object is an object that is disposed to give a certain kind of observer a certain kind of experience under certain illumination conditions.<sup>127</sup> Unfortunately, it has also proven difficult to find any one-to-one correlations between kinds of surfaces + illumination conditions and certain colour experiences in certain perceivers. The same coloured surface under the same illumination conditions can appear differently to different observers or even to the same observer at different times. Furthermore it is difficult to define the relevant kind of observer. In any case, being a human being with normal colour vision cannot be that kind. Two people with normal colour vision

---

<sup>126</sup> See Hardin 1988 for a survey of this problem.

<sup>127</sup> The theory of colours as dispositional properties has been put forward by e.g., J.C.C. Smart and David Armstrong. It is outlined in e.g., Armstrong & Malcolm 1984.

may look at the same object at the same time and yet have different colour experiences. One sees the surface as e.g., bluish green and the other as uniquely green.<sup>128</sup> These facts make dispositional characterisations of colour difficult.

Lycan claims that the colours we see are phenomenal and that they are “parasitic on the real colours of physical objects.”<sup>129</sup> Physical colours, on the other hand are regarded as dispositional properties of physical objects.

“It construes physical colors as physical properties of objects, but only as very modest ones; it does not flatter those properties in any way. They are roughly the properties that would, indeed do, constitute their owner’s dispositions to produce the corresponding sensations in sentient observers under normal viewing conditions, and as we know, these properties are an unruly, rough, and ragged lot. Almost certainly they form no natural kind.”<sup>130</sup>

The problem with Lycan’s characterisation is that it threatens to become circular, physical colours are picked out with reference to the human visual system, while phenomenal colour is explicated in terms of what kind of experiences physical colour produces in sentient beings like us. A red object is an object that produces red experiences in creatures of our kind under normal viewing conditions, and “normal viewing conditions”<sup>131</sup> are the conditions under which an object of a certain colour produces the corresponding experience in normal viewers. Lycan claims that these problems can be worked around since “normal viewing conditions” can be specified in a way that does not make use of perceiver’s colour experiences.

“Here I turn to the work of Roger Shepard (1990, 1991, 1992), who offers an independent characterization of ‘normal viewing conditions’ in evolutionary/ecological terms. The idea is roughly that normal conditions are those that prevailed on the earth’s surface throughout the eons during which color constancy was established in our ancestors, but this is a longish story.”<sup>132</sup>

---

<sup>128</sup> Hardin 1988, pp. 89-91,

<sup>129</sup> Lycan 1996, p. 72.

<sup>130</sup> Ibid., pp. 72-73.

<sup>131</sup> Ibid., p. 74.

<sup>132</sup> Ibid.

So, normal viewing conditions, standard human physiology<sup>133</sup>, and verbal reports about colours of objects, yields a “reference-fixing triangulation of any given Armstrongian color property”, according to Lycan.<sup>134</sup>

Shoemaker has another idea of what our experiences represent. He differentiates between, on the one hand, physical colours, and on the other, phenomenal colours. While physical properties are objective properties that the objects have independently of any viewer, phenomenal properties are dependent on the viewer, but are still properties of physical objects. The phenomenal properties, which Locke called “secondary properties”; smell, colour, taste and so on, are properties of physical objects and not properties of our experiences, Shoemaker says.<sup>135</sup> Shoemaker, unlike Lycan, claims that our perceptual experiences represent the phenomenal properties of external objects and not only their physical, objective properties.

Shoemaker argues that it would be preferable if we could locate the phenomenal properties to where they seem to be, i.e., with the physical objects. The question is: how can colours be physical properties of objects if the environment is anything like physical science tells us? If we put colours back into our experiences, Shoemaker says, the account will accommodate the way the world is described by science, but then again, locating them there seems “to fly in the face of phenomenology” since we do certainly not experience them as parts of our experiences, but as properties of the perceived objects.

Shoemaker argues that the possibility of spectrum inversion is compatible with the view of phenomenal properties as properties of objects. Here is the main line of his argument: Suppose Jack and Jill are watching the same tomato. They are both “normal” perceivers and the lighting conditions are normal. To Jack the tomato has the same colour as cucumbers normally have to Jill and to Jill the tomato looks what most of us think of as red. Let Q1 be the quale associated with the phrase “tomato colour” in Jack and Q2 the quale associated with “tomato colour” in Jill. Jack’s experience represents the tomato as having one property, Jill’s represents it as having another.

---

<sup>133</sup> It seems to be a fact that there is no “standard human physiology” either. See Hardin 1988.

<sup>134</sup> Lycan 1996, p. 74 .

<sup>135</sup> This idea should also imply that such claims can be made about “primary properties” as well, Shoemaker says. But, he continues, in the case of the secondary properties the idea derives much of its plausibility from the fact that there is an “explanatory gap” between how these properties are represented by our experiences and what we know about them in other ways. I am not sure whether he thinks that this gives us reason to believe that only the secondary properties are “phenomenal” in the sense he accounts for.

Now, does this mean that the tomato has different properties at the same time, or that either Jack or Jill (or both) misperceive? It means that the tomato has different *phenomenal properties*, Shoemaker claims.

“Assuming the possibility of spectrum inversion, this would mean that the properties should be of a kind such that different perceivers can, under the same objective conditions, perceive the same objective thing, or things of exactly the same color, to have different properties of that kind, and perceive things having different colors to have the same properties of this kind, this because of differences in their subjective constitutions.”<sup>136</sup>

The solution to the problem lies, according to Shoemaker, in maintaining that the phenomenal properties of the tomato are relational. This opens for the possibility of Jack and Jill to have different qualia representing the same objective colour, Shoemaker claims. Jack’s experience correctly represents the tomato as having one phenomenal property, and Jill’s experience correctly represents the tomato as having another. This, however, entails that the phenomenal properties of the tomato are not the same as its objective properties.

One objection, raised by Shoemaker himself, is that *it does not seem to us* as if the colours we perceive are relational properties. Just as he takes the phenomenological fact seriously that phenomenal properties seem to be properties of objects, should he not also pay attention to the fact that they seem to be monadic properties of objects and most certainly not *relational* ones? But, in this case, Shoemaker does not take the way properties seem to us to be a good guide to their real status. He brings up some other examples where properties *seem* monadic, but *are* relational. Being heavy is really a relational property, even if perceived as being monadic. A football is heavy to a 1-year-old child, but it is not heavy to a teenager. And, he continues, the relation “to the right of”, is really a triadic one (a relation between two objects and one perceiver), while we tend to regard it as dyadic.

Importantly, Shoemaker argues that our perceptual states primarily represent the phenomenal properties. In 1994 Shoemaker says:

“If I am right, a color experience represents an object as having a ‘phenomenal’ property that is constituted by a relation to sense-experience – and it is the representation of this property that gives the experience its phenomenal character. The experience *also* represents the object as having

---

<sup>136</sup> Shoemaker 1996c, p. 253.

a certain color. [...] [T]he experience represents the color *by* representing the phenomenal property.”<sup>137</sup>

In a later discussion he says:

“If the phenomenal character of perceptual experiences consisted in their representational content, and if this content consisted entirely in the representation of ‘objective’ properties, then experiences that have the same objective representational content would be alike in phenomenal character, whether the perceptual mechanisms of the subjects of these experiences were the same or different.”<sup>138</sup>

Shoemaker uses a distinction between looking like something in a *phenomenal* sense and looking like something in a *doxastic* sense, named thus by Dretske.<sup>139</sup> A tomato may look purple to an observer while she does not believe that it actually is purple (that it has the objective property of purple). This is to look purple in a phenomenal sense, to look<sub>p</sub> purple. Or the tomato looks red to an observer in the sense that she perceptually believes that it is red. It looks doxastically red (looks<sub>d</sub> red). Shoemaker’s motive for using this distinction is to show that perceiving a colour that the physical object does not have (objectively) is not necessarily a case of misperception. It is only, Shoemaker claims, when a thing appears doxastically to have a property, which it does not have, that veridicality fails.

“[T]he transition from the ways things appear phenomenally to what objective properties they have, and what objective properties they appear to have, rests on certain contingencies. These include facts about illumination conditions and the spatial relation of the perceiver to the object, and facts about how these combine with the objective properties of a thing to determine how it will look<sub>p</sub>.”<sup>140</sup>

So, according to this view, our (scientific) knowledge about illumination and so forth affects the way we perceive<sub>d</sub> things. If we have reason to believe that the object perceived does not really have the colour it appears<sub>p</sub> to have, it will appear<sub>d</sub> not to have this colour. In sum, Shoemaker argues that our

---

<sup>137</sup> Shoemaker 1994, p. 35.

<sup>138</sup> Shoemaker 2002 p. 458.

<sup>139</sup> Dretske 1995.

<sup>140</sup> Shoemaker 2002 p. 461.

perceptual states represent both objective and phenomenal properties of experiences.

The distinction between phenomenal colours and objective colours made by Shoemaker has been criticised by e.g., Tye. Tye claims that the implication that colours are not actually perceived “draws a veil over the colors.”

“Drawing this veil is tantamount to erecting an appearance/reality distinction for the colors themselves. The coherence of such a distinction is dubious at best[...]<sup>141</sup>

Uriah Kriegel defends the distinction against Tye’s criticism.<sup>142</sup> According to colour objectivism, which Tye himself endorses, there are objective mind-independent facts about colour. If there are such facts, a subject might get them wrong and when she does, things will appear to have colours they in reality do not have, Kriegel says. But we can give Tye’s objection another interpretation, which Kriegel also notes. Namely that it follows from Shoemaker’s view, that objective colours *cannot be known* at all. If our knowledge of colour is restricted to phenomenal colours, what reasons do we have for postulating objective colours in the first place? As Kriegel accurately points out, even if it is the case that we cannot perceptually experience objective colours, we can still know about them in some other way. There are cases, as the one with the purple-looking tomato, where we have reasons to believe that the colour experienced is not the real colour of an object.

So, while Lycan, Tye and Dretske all argue that our experiences represent objective properties of physical objects, Shoemaker claims that they (primarily) represent phenomenal properties. We must thus keep in mind that while a “phenomenal property” according to the first trio (with the possible exception of Lycan) is the same as a quale, it is something else in Shoemaker’s opinion. I will leave these problems here and turn to a related problem that the representational account of qualia brings about.

From time to time when we see some object, there does not seem to be anything in the real world corresponding to this experience. I may dream of a red rose, I may hallucinate a red rose or I may mistake a red crumpled napkin for a red rose. Qualia in this case are not represented properties of

---

<sup>141</sup> Tye 2000a, p. 4.

<sup>142</sup> Kriegel 2002.

any physical object. Both Dretske and Tye hold that in the event of a hallucination, there is no object to be aware of, only properties. Tye writes:

“There is, then, a definite content to Paul’s hallucinatory experience. But there is no object, mental or otherwise, that Paul hallucinates.”<sup>143</sup>

Lycan has a similar idea, he speaks of intentional inexistents.

“Of course it is characteristic of intentional contents that they may or may not actually exist; and that is how we evade the dilemma. Your visual system quite often portrays, alleges something green. But, vision being not entirely reliable, on a given occasion the green thing may not actually exist.”<sup>144</sup>

The best semantics for intentional inexistents is a possible-world semantics, Lycan claims. A green after-image is a representation of a green patch existing in a possible, alternative world. To see a green after-image is to perceive *as if* there was something green before me.

There are certain problems with this account though, how does Lycan propose that we decide who the “as if” is as if for? Is it as if for all human beings, for human beings with “normal” colour vision, or as if for any possible observer? A colour blind person would experience what a “normal” person would consider green, if a red object was before him. Bernard Kobes presents several counterexample to Lycan’s account of “as if”.<sup>145</sup> In his reply to Kobes, Lycan presents his own version of one of these counterexamples (which I prefer since it makes the same point with less fanciful components). Lycan invites us to think of a population, the “Antigreens” to whom green physical objects are invisible. A green patch would hence be invisible, if sensed *as if* a green object was present to an Antigreen. So the meaning of “as if green” seems to be dependent on what kind of perceiver one has in mind. Lycan tries to elaborate his account in order to avoid this conclusion:

“Given any organism of any species anywhere, what it is for that organism to have a green sensation is for the organism to be sensing as *we* (normal human beings) would be sensing *in the actual world* were a green object present to us. [...] the Antigreens [...] do not have green sensations when confronted by green objects, precisely because they do not sense *as we do*

---

<sup>143</sup> Tye, 2002, pp. 448-449.

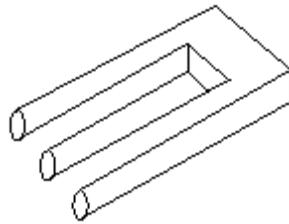
<sup>144</sup> Lycan 2001c, p. 18.

<sup>145</sup> Kobes 1991, pp. 154-155.

in such circumstances. This move makes ‘green sensation’ into a rigid designator whose reference is fixed by allusion to ourselves in the actual world (which is fine in itself; the word ‘green’ is our word, after all).”<sup>146</sup>

“As if green” is hence as green would be like to a normal human being who lives in this actual world.

Another, slightly more serious problem, is that we seem to be able to see things that do not exist in any possible world. Look at this drawing of the so-called “Devil’s fork”.<sup>147</sup>



It looks as a drawing of an object that has both two and three “arms” at the same time. If we follow Lycan’s suggestion, the fork we see would be an intentional inexistents, but in this case it does not exist in any possible world.

#### 4.2.2 How do our experiences represent properties of objects?

Another problem for the representationalist theory is to explain in virtue of what an experience represents the properties it does. For an experience to represent the world as being a certain way, it seems as if there must be a set of semantic conditions under which it does this correctly, truthfully, or the like. Does it follow that representational content must be conceptually structured? A positive answer to that question seems to entail e.g., that I have one concept for each shade of green I can discern in the spring landscape, and one concept for every nuance of the music which I can hear being played on the radio right now. If we do have all these concepts, we are obviously not able to verbalise all of them. There are (at least) two ways to account for this, given the assumed accuracy of representationalism. One could claim that we do have concepts for all shades of colour and so on, but

---

<sup>146</sup> Lycan 1996, p. 86.

<sup>147</sup> Oscar Reutersvärd invented a similar thing in 1958.

that those concepts are somewhat different from “standard concepts”, or one could argue that perceptual experiences have “non-conceptual content”.

The notion of nonconceptual content was explicitly introduced by Gareth Evans.<sup>148</sup> Christopher Peacocke, another influential proponent of non-conceptual content, claims that the fineness of grain of perceptual experience outstrips the conceptual capacities of the perceiver, but that perceptual states can be described in terms of “scenario content”.

“I suggest that one basic form of representational content should be individuated by specifying which ways of filling out the space around the perceiver are consistent with the representational content’s being correct. The idea is that the content involves a spatial *type*, the type being that under which fall precisely those ways of filling the space around the subject which are consistent with the correctness of the content. On this model, correctness of a content is then a matter of instantiation: the instantiation by the real world around the perceiver of the spatial type which gives the representational content in question.”<sup>149</sup>

Opponents of the idea that perceptual content is non-conceptual have argued that the “fineness of grain” of perceptual experiences can in fact be accommodated at the conceptual level. John McDowell, for example, suggests that the conceptual content of perceptual experiences is given by demonstrative concepts, such as *that shade* or *that sound*.<sup>150</sup> The information lost in the transition from perception to perceptual belief does not entail that there are two different types of content in play, but should rather be understood as a shift from a more determinate to a less determinate type of conceptual content, McDowell argues.

The philosophers whose theories are at issue here all argue in favour of some kind of non-conceptual content. For Dretske and Tye this kind of content plays an important role in their theories of consciousness. States with nonconceptual or “systemic” or “analogue” content play a certain role in the cognitive system and that is what makes them phenomenally conscious. Dretske writes:

“[E]xperiences are to be identified with states whose representational properties are *systemic*. Thought (conceptual states in general), on the other hand, are states whose representational properties are *acquired*. As a result,

---

<sup>148</sup> Evans 1982.

<sup>149</sup> Peacocke 1992, p. 105.

<sup>150</sup> McDowell 1994.

experiences have their representational content fixed by the biological functions of the sensory systems of which they are states. How an experience represents, [systemically] the world is fixed by the functions of the system of which it is a state. The quality of a sensory state – how things look, sound and, feel at the more basic (phenomenal) level – is thus determined phylogenetically.”<sup>151</sup>

#### 4.3 HIGHER ORDER THEORIES

There are different versions of the higher order approach to consciousness and introspection. The HO theories claim that a mental state, M is (phenomenally) conscious if (and only if) the subject is *conscious of* M, by means of having a higher order representation of M. A first order mental state, FM, represents some state of affairs in the world, but it does not necessarily represent this *to* the subject. A subject can according to these theories have a perceptual experience and the experience is like something in itself, but it is only when the subject becomes aware of that perceptual experience that it will be like something *for the subject*.<sup>152</sup>

---

<sup>151</sup> Dretske 1995, p. 15.

<sup>152</sup> There are other representationalist accounts of introspection and phenomenal consciousness which we will not consider here. There are e.g., some theories that are similar to the HO theories insofar as they are built on representationalism and argue that phenomenal consciousness requires awareness of mental states, but according to these versions the higher order state is an intrinsic aspect of the first order state. Proponents of this view are, besides Brentano, e.g., Rocco Gennaro (1996, 2004), Uriah Kriegel (2003) and Thomas Natsoulas (1999). Gennaro calls this the “wide intrinsicity view”. Robert Van Gulick (2004) also launches a variant of this theory. He claims that the HO state is part of the “global conscious state” and labels such states “HOGS” (higher-order global states). “The meta-intentional content is carried not by a distinct and separate vehicle but rather a complex global state that includes the object state as a component.” Van Gulick has however a functionalist view of mental content, “a state’s content is a function of the functional role within the system of which it is part.” (2004, p 81) There is one important objection to this account. It seems as if the secondary awareness of the act and the primary awareness must be distinct states since they apparently have distinct truth conditions. (This critic is put forward by e.g., Rosenthal 1997, pp 746, and Amie Thomasson 2000, p 199). As Thomasson puts it “the (first-order) thought that this is an orange tree has different truth conditions from the thought that I think that this is an orange tree, and so, insofar as truth conditions are relevant to the content of an act, the acts must have different contents, and be distinct mental acts.” (p 199). Brentano acknowledges that if we individuate presentations by the “number and variety of objects” presented we would have different presentations. Instead he suggests that presentations

### 4.3.1 Higher order perception

Exactly what is it that constitutes the difference between being and not being conscious of our own perceptual states (or first order states in general)? Armstrong tries to illustrate this difference by describing a phenomenon that may occur during long distance truck (or car) driving. This example is henceforth widely used by all the HO theorists.

“After driving for long periods of time, particularly at night, it is possible to ‘come to’ and realize that for some time past one has been driving without being aware of what one has been doing. The coming-to is an alarming experience. It is natural to describe what went on before one came to by saying that during that time one lacked consciousness. Yet it seems clear that, in the two senses of the word that we have so far isolated, consciousness was present. There was mental activity, and as part of that mental activity, there was perception. [...]. What is it that the long-distance truck driver lacks? I think it is an additional form of perception, or, a little more cautiously, it is something that resembles perception. But unlike *sense*-perception, it is not directed toward our current environment and/or our current bodily state. It is perception of the mental. Such ‘inner’ perception is traditionally called introspection, or introspective awareness.”<sup>153</sup>

Being aware of just the road and the traffic is not sufficient for being aware of what one is doing. That seems about right. Quite often, perhaps even most of the time when we are driving we are aware of the traffic and so on, but we not aware of us *driving* the car. That is, we are not paying attention to the fact that it is we who are driving the car. Armstrong claims that the driver lacks “introspective awareness”, but why is it that the driver will become aware of what she is doing by “perceiving” her own perceptual states?

Carruthers describes the difference between being conscious and not being conscious of our perceptual states as a difference between being conscious of “what the world is like for the organism, and being conscious

---

should be individuated by “number of mental acts” and he insists upon there only being one such act. Brentano stresses that inner awareness is not observational, because if it were it would require a different act doing this observing, but he is not clear on the point on how we should understand the act’s awareness of itself.

<sup>153</sup> Armstrong 1997, pp. 723-724.

of what the organism's experience of the world is like for the organism".<sup>154</sup> It is not only that we become aware of the *fact* that we are perceiving the world, we become aware of what our perceptual experiences of the world are like. This is probably what Armstrong is driving at, as well. According to Armstrong the subject needs to be introspectively aware of her mental states in order to be conscious of what she is doing.

“In sense-perception we become aware of current happenings in the physical world. A perception is therefore a mental event having as its (intentional) object situations in the physical world. In introspection, on the contrary, we become aware of current happenings in our own mind. Introspection is therefore a mental event having as its (intentional) object other mental happenings that form part of the same mind.”<sup>155</sup>

According to the philosophers now under discussion, the mechanism that enables introspective awareness is one which in many respects resembles that of external perception. Paul Churchland, another proponent of the HOP theory, writes:

“[O]ne's introspective consciousness of oneself appears very similar to one's perceptual consciousness of the external world. The difference is that, in the former case, whatever mechanisms of discrimination are at work are keyed to internal circumstances instead of external ones”<sup>156</sup>

Lycan ascribes to the same basic view. The perceptual or attentional mechanism by which we attend to our mental states is *probably* a mechanism in the brain:

“Something cognitive, and presumably something neurophysiological, subserves this ability.”<sup>157</sup>

Lycan further differentiates between just *perceiving* one's own perceptual states and *attending* to them. In subsection 3.2.2 we called these processes: *NAHOP* for “non-attentive higher order perception”, and *AHOP* for “attentive higher order perception”. *AHOP* is hence introspection.

---

<sup>154</sup> Carruthers 2000a, p. 129.

<sup>155</sup> Armstrong 1968, p. 323.

<sup>156</sup> Churchland 1988, p. 74.

<sup>157</sup> Lycan 1996, p. 17.

“[I]ntrospection is the operation of an internal attention mechanism that monitors experiences and produces second-order representations of their properties, including especially their feature-representational properties.”<sup>158</sup>

To be phenomenally conscious of something – a red tomato, say – is to have a NAHOP of an experience that represents a red tomato. To introspect a perception of a tomato, is to have an AHOP of an experience that represents a red tomato.

The HOP theory can be illustrated thus:

A lower order perceptual experience:

$\epsilon \rightarrow \text{🍏}$

A higher order non-attentive perceptual experience (NAHOP):

$\epsilon \rightarrow \epsilon \rightarrow \text{🍏}$

A higher order attentive experience (AHOP):

$\epsilon \Rightarrow \epsilon \rightarrow \text{🍏}$

$\rightarrow$  means: non-attentively perceives

$\Rightarrow$  means: attentively perceives

$\text{🍏}$  means: red apple

$\epsilon$  means: perceptual experience

$\epsilon$  means: higher order perception

### 4.3.2 Higher order thinking

David Rosenthal has long argued in favour of a different version of the higher order model of consciousness, a version that conceives of the higher order representational relation in terms of thinking instead of perceiving.<sup>159</sup>

“The two ways of being conscious of things are sensing them and having thoughts about them as being present. Since we are not conscious of our mental states by sensing them, the best explanation of how we are conscious of some of our mental states is that we have higher-order thoughts (HOTs) about them. [...] It would be explanatorily empty to

---

<sup>158</sup> Lycan 2003 p. 26.

<sup>159</sup> Varieties of the HOT theory have also been put forward by e.g., Mellor 1977-78, Rolls (1998), Kobes 1996, Gennaro 1996.

insist that we are conscious of them in some third way unless we have an independent grasp of what that third way consists in.”<sup>160</sup>

The HOT theory, just like its HOP “cousin” postulates the existence of mental states of at least two orders. The HOT theory entails that qualia are intrinsic non-representational properties of perceptual states and that the subject has to be conscious of those qualitative properties in order to be phenomenally conscious.

“[T]he theory does argue that there is something qualitative that it’s like for one to be in a state whenever one has a suitable HOT that one is in that state and the HOT characterizes the state in qualitative terms. If so, having such a HOT does constitute there being something qualitative that it’s like for one to be in the state. The theory claims not that HOTs cause qualitative consciousness to occur, but that the having of suitable HOTs is what it is for qualitative states to be conscious.”<sup>161</sup>

What kind of thoughts are higher order thoughts, according to Rosenthal? Will any propositional (non-perceptual) state, regardless of mental attitude, make another state conscious, as long as it is *about* that state? Higher order thoughts are affirmative thoughts to the effect that a subject is in a particular state, Rosenthal claims. Furthermore, they will have to be occurrent thoughts. Just being disposed to have a thought about something does not by itself make us conscious of anything, Rosenthal claims.

While the HOP theory pictures introspective awareness as attentive higher order perception, Rosenthal gives another account of introspection. He thinks that introspection entails a third order of representations.

“The HOTs in virtue of which we are conscious of our mental states in ordinary, nonintrospective cases are not, themselves, conscious thoughts; HOTs make one aware of various mental states, but without one’s being conscious also of the HOTs themselves. When one introspects a state, one deliberately focuses attention on it. One thereby becomes aware not only of the introspected state, but also of one’s being conscious of it. So in these cases the relevant HOTs are themselves conscious thoughts.”<sup>162</sup>

---

<sup>160</sup> Rosenthal 2002a, p. 719.

<sup>161</sup> Rosenthal 2004, p. 39.

<sup>162</sup> Rosenthal 2002a, p. 719.

The HOT theory can thus be illustrated thus:

A first order perceptual experience:  $e \rightarrow \text{apple}$   
A second order thought:  $\varepsilon \rightarrow e \rightarrow \text{apple}$   
A third order thought (introspection):  $\epsilon \rightarrow \varepsilon \rightarrow e \rightarrow \text{apple}$

$\rightarrow$  means: perceives  
 $\rightarrow$  means: thinks she is in  
 $\text{apple}$  means: red apple  
 $e$  means: perceptual experience  
 $\varepsilon$  means: second order thought  
 $\epsilon$  means: third order thought

### 4.3.3 Higher order dispositional thinking

A third version of the higher order representational theory is recommended by Peter Carruthers. This theory claims that in order for a mental state of the first order to be conscious, it needs only to be *available* for a higher order thought. It is sufficient for a first order experience to be disposed to cause a higher-order belief that the subject is having it, for the subject to be conscious of this experience.

“Any occurrent mental state M, of mine, is conscious = M is disposed to cause an activated belief (possibly an non-conscious one) that I have M, and to cause it non-inferentially. In contrast with the actualist form of HOT theory, the HOTs which render M conscious are not necessarily actual, but potential.”<sup>163</sup>

Carruthers presents a model for how this works. Conscious experiences occur, according to him, when perceptual contents are fed into a special short-term memory buffer, whose function it is to make these contents available to cause higher order thoughts about themselves.

“Perceptual contents are regularly passed to two or more short-term memory stores, C (conscious) and N (non-conscious), to be integrated with the subject’s goals in the control of action. But C itself is now defined, *inter alia*, by its relation to HOTs – any of the contents of C being

---

<sup>163</sup> Carruthers 2000a, p. 227.

*apt* to give rise to a HOT about itself, should circumstances (and what is going on elsewhere in the system) demand.”<sup>164</sup>

Carruthers claims that first-order representations of the environment will take on a “dual content” when they are disposed to cause the appropriate HOTs.

“Each experience of the world-body becomes at the same time a representation that just such an experience is taking place; each experience with the content *red<sub>s</sub>*, say is at the same time an event with the content *seems red<sub>a</sub>* or *experience of red<sub>s</sub>*. And [these experiences] have these contents categorically, by virtue of the *powers* of the HOT consumer system, in advance of any HOT actually being tokened.”<sup>165</sup>

One might think that according to Carruthers a perceptual experience of e.g., a black cat becomes a conscious experience when it actually gives rise to a HOT, but that is not so, since it would make his theory collapse into the HOT theory. My experience of a black cat is already conscious when it is disposed to cause a higher order belief that I have this experience. It is so disposed if it has the dual content (of e.g.,) [black] and [experience of black].

A conscious experience could perhaps be illustrated thus:

[ $\epsilon \rightarrow \text{apple}$ ] is disposed to cause  $\epsilon$ .

#### 4.4 ONE LEVEL THEORIES OF CONSCIOUSNESS

The “one-level theorists” argue that phenomenal consciousness can be reduced to a certain kind of representational content (systemic, or non-conceptual). These theories maintain that “what-it-is-like for the subject” or “the subjective dimension” of experiencing does not derive from the subject’s being aware *of* the experience as such. While the higher order theorists assume that awareness of what it is like requires awareness of perceptual experiences, the one level theorists maintain that to be aware of what it is like to perceive, is to be aware of what the external objects are like.

---

<sup>164</sup> Ibid., p. 228.

<sup>165</sup> Ibid., p. 242.

Both Dretske and Tye are in favour of strong representationalism. Tye brings forward the “transparency phenomenon of perceptual experiences”, originally described by Moore<sup>166</sup> and uses this as an argument for his strong representationalist view. Experiences are transparent insofar as when you try to attend to or focus on the experience, you always end up (just) being aware of what the experience is an experience of. This shows that the phenomenal character you are aware of when introspecting is nothing but the intentional content of your perceptual state, Tye claims.

#### 4.4.1 Dretske’s theory of introspection

Dretske’s and Tye’s theories of introspection are similar, but Dretske has elaborated his a bit further. In 1995 Dretske writes:

“What one comes to know by introspection are, to be sure, facts about one’s mental life – thus (on a representational theory) representational facts. These facts are facts, if you will, about internal representations. The objects and events one perceives to learn these facts, however, are seldom internal and never mental [...] One becomes aware of representational facts by an awareness of physical objects.”<sup>167</sup>

In the same book Dretske offers a model of introspection that is based on a perceptual analogy, which he calls “displaced perception”. We come to know that something is the case, as e.g., that the postman has arrived, either by hearing or seeing *the postman*, or by hearing or seeing *something else*, like a dog barking. Introspective knowledge can be accounted for in a way similar to the latter case.

“[A]n experience (of blue, say) is conceptually represented as an experience of blue via a sensory representation not of the experience, but of some other object. One comes to know (the fact) that one is experiencing blue by experiencing, not the experience of blue, but some displaced object [...] [T]his displaced object is (typically) the object the experience of blue is an experience of – i.e., the blue object one sees.”<sup>168</sup>

We are aware of how our experiences are, not by directly attending to the experience as such, but by attending to what the experience is an experience of. Introspection is, thus, an indirect affair, just as displaced perception is.

---

<sup>166</sup> Moore 1922, p. 22.

<sup>167</sup> Dretske 1995, p. 40.

<sup>168</sup> Ibid., p. 44.

Dretske elaborates his ideas of introspection in a later paper.<sup>169</sup> The paper opens with two propositions. When we try to assert both, we inevitably run into problems, he points out. These propositions are: (1) perceptual experiences are inside a person, and (2) nothing inside a person has (or needs to have) the properties we are conscious of when we have these experiences.<sup>170</sup> Dretske's point is that the experience of for example a pumpkin is not intrinsically orange and bulgy, and we can hence not come to know the content of the pumpkin experience by studying the experience-qua-mental object.

Dretske sets out to specify how the indirect knowledge in introspection works. It is quite possible, he says, to be aware of facts about physical objects without being directly aware of either the object as such or *its properties*. He thus distinguishes between: *object-awareness*, *property-awareness* and *fact-awareness* (*o*-awareness, *p*-awareness and *f*-awareness, respectively). Let *o* be an object, an event or a state (a spatio-temporal particular) and *P* a property of *o*. We can be aware, both of *P*, of *o* and of the fact that *o* is *P*. These forms of being aware are (conceptually, not causally) independent of one another. The phenomenal character of perceptual experiences is determined by the qualities one is *p*-aware of. *Object-awareness* has to do with the causal relations experiences have to objects in the world and *fact-awareness* has to do with what knowledge our perceptual experiences give rise to. We can, thus, be aware of an object while not being aware of its properties and we can be aware of some properties that an object has while not at the same time being aware of the object. Dretske claims that being aware of a hallucinatory object constitutes an example of property-awareness without object-awareness. We will return to this claim below.

When we for example study the minute hand of a clock under the mistaken assumption that the clock is broken, we are aware of the object, (the hand; *o*), but neither of its movement *R* nor of the fact that it moves (that *o* is *R*). Or, assume that I watch the same hand for several minutes and at some point realise that the hand's position has changed. In this case I am aware that it has moved, but I was not aware of the movement as such. I am aware of *o* and that *o* is *R*, but not of *R* itself. We are also sometimes aware of the fact that an object has certain properties without being aware of either the object or its properties. Dretske uses knowledge of electrons as an

---

<sup>169</sup> Dretske 1999.

<sup>170</sup> Dretske uses "inside" in the spatial sense here. Perceptual experiences are inside people's heads.

example. Electrons are objects with certain properties such as spin. We can be aware of the fact that electrons have spin, without being directly aware either of the electrons or of the spin. This is then an example of *f*-awareness without *p*-awareness or *o*-awareness.

Dretske introduces a notation for experiences and their properties. “*e*” stands for an experience and “*P*” for a property of an experience. Dretske argues that the same circumstances that prevailed in the case of perception hold here. We can be aware that *e* is *P*, without being (directly) aware of either *e* or *P*.

“The mistake in traditional arguments lies in failing to distinguish between *f*-awareness of experience, that it has phenomenal character *P*, on the one hand, on the other, *p*-awareness of the qualities (e.g., *P*) that give it this character. Failing to distinguish these forms of awareness, one concludes, mistakenly that awareness of what it is like to see (experience) pumpkins must be awareness of the properties (i.e., *P*) of these experiences.”<sup>171</sup>

The analogy with the different kinds of perceptual awareness shows that we do not have to be aware of the experience-qua-object nor of its properties in order to be aware of the fact that it has certain properties, Dretske argues.

The question is: what other objects and properties are we aware of in order to become indirectly aware of facts about our own experiences? The answer must be: the objects and properties these experiences make us aware of.

“[T]he qualities one becomes *p*-aware of in having a perceptual experience are qualities of external objects (the pumpkins) that one experiences, not qualities of the pumpkin-experience. One becomes *f*-aware of experience – that it is *P* – by *p*-awareness of *P* – the pumpkin’s properties. The reason *p*-awareness of *P* can make one *f*-aware that one’s experience is *P* is that *P* is the property of being an experience, in fact a *p*-awareness, of *P*.”<sup>172</sup>

This account is compatible, Dretske says, with the statement that experiences are inside our heads, and that they do not have the properties we are conscious of when we have these experiences.

---

<sup>171</sup> Dretske 1999, pp. 110-111.

<sup>172</sup> Ibid., p. 112.

#### 4.4.2 Tye

Tye, like Dretske, claims that conscious experiences have a certain kind of content that distinguishes them from the non-conscious ones, since it makes them available for behavioural control. Introspective awareness is also here taken to be “fact-awareness”.

“When we introspect our experiences and feelings, we become aware of what it is like for us to undergo them. But we are not directly aware of those experiences and feelings; nor are we directly aware of any of their qualities. The qualities to which we have direct access are the external ones, the qualities that, if they are qualities of anything, are qualities of external things.”<sup>173</sup>

Introspective awareness is awareness that an experience with a certain phenomenal character is present, Tye claims. Both Tye and Dretske add that this account fits nicely with the linguistic constructions that we use when we talk about introspective awareness.<sup>174</sup> Tye writes:

“To talk of our being aware of the phenomenal character of an experience or of how an experience ‘feels’ is to use a generic perceptual verb (‘aware of’) followed by an abstract noun (‘the phenomenal character’) or an interrogative nominal (‘how the experience feels’). In cases of this sort, where there is a perceptual verb, the abstract noun or interrogative nominal typically stands in for a factive clause so that what is being described is (a species) of awareness of some fact.”<sup>175</sup>

The most natural interpretation of “being aware of the phenomenal character” is according to Tye, to be aware of a fact and not to be aware of an object or some property of an object. The fact-awareness analysis of introspective awareness does not distort the meaning of the original phrase, Tye maintains.

According to Tye, introspection of phenomenal character is a *reliable automatic process*. The process takes the subject from being in one state to being in another state. While the initial state can be described as an *awareness of something*, the content of the second state is *awareness that*. The first state can hence be S’s awareness of a cup of coffee and the second state: S’s awareness that she has an experience of a cup of coffee. The first state has

---

<sup>173</sup> Tye 2002, p. 144.

<sup>174</sup> Dretske 1993.

<sup>175</sup> Tye 2002, p 145.

non-conceptual content in virtue of being an experience without belief, while the latter has conceptual content on account of being a regular belief.

#### 4.4.3 Shoemaker

Sydney Shoemaker has long argued that introspective awareness is awareness of facts about our experiences.

“Introspective awareness is awareness *that*. One is introspectively aware that one has an experience with a certain representational content, and with the phenomenal character that involves.”<sup>176</sup>

Shoemaker argues that introspective beliefs are not justified by some other beliefs or experiences the subject may have. The content of an introspective state simply *embeds* the content of a perceptual state.

“The content of this awareness [the introspective one] will embed the content of the perceptual state, in the same way that the content of an introspective belief about visual experience will embed the content of the visual experience.”<sup>177</sup>

In order to properly understand Shoemaker’s comprehension of introspective awareness of perceptual states, one must first be familiar with his theory of phenomenal properties which was briefly described in subsection 4.2.1. He distinguishes between the objective properties of a physical object and its phenomenal properties. There are three kinds of the latter sort; *occurrent appearance properties*, *dispositional appearance properties* and *higher order dispositional appearance properties*. *Occurrent appearance properties* are properties of actually causing an experience of a certain sort in a perceiver. *Dispositional appearance properties* are properties of being “apt to produce experiences of a certain sort in some kinds of observers when those observers are related to it in a certain way”.<sup>178</sup> One object can have a vast number of dispositional appearance properties, which entails that it may be apt to appear differently to different observers due to differences in their perceptual systems.<sup>179</sup> A *higher order dispositional appearance* property is a

---

<sup>176</sup> Shoemaker 1996c, p. 255.

<sup>177</sup> Shoemaker 2000, p. 464.

<sup>178</sup> Shoemaker 2002, p. 460.

<sup>179</sup> It is not clear to me whether Shoemaker thinks that the reason the same object can appear to have different colours at different times to the same observer, is also in virtue of its

property “shared by all things that are disposed, under some circumstances or other, to appear in a certain way to normal observers of one or more sorts situated in one or another way with respect to them.”<sup>180</sup>

This “tristinction” may be exemplified thus: assume that I am standing by the pasture and watch my horses. I will (hopefully) perceive a white horse. The horse, then, has an occurrent as well as a dispositional appearance property of looking white to my kind of observer. But the horse is at the same time disposed to appear differently coloured depending on illumination conditions and the perceptual makeup of the observer. The horse is disposed to appear bluish in moonlight and golden at sunset and it might be disposed to look \*red\* to an alien. Furthermore, if the horse looks white to me and snow looks white to me, my experience represents the horse and snow as having the same occurrent appearance properties and the same higher order appearance dispositional property, but the dispositional appearance property of the horse differs from that of the snow, Shoemaker seems to argue.

A perceptual experience represents, among other things, instantiations of occurrent appearance properties, which means, Shoemaker says, that it represents states of affairs that *are partly mental*. The instantiation of such a property is a mental state of affair, since it requires that the thing having the property is actually perceived by someone. And this, in turn, means, according to Shoemaker, that the contents of the perceptual state and the introspective state are not logically independent of each other (which they would be if the properties perceptual states represent were objective properties of objects), but instead “closely related”.<sup>181</sup>

“The relation of ‘That looks<sub>p</sub> blue to me’ to ‘I am having an experience as of something that looks<sub>p</sub> blue’ seems to be one of conceptual entailment. So the introspective awareness expressed by the second proposition seems implicit in the perceptual awareness expressed by the first.”<sup>182</sup>

---

dispositional appearance properties. He writes: “For it will be disposed to appear in different ways depending on the illumination conditions, the distance of the observer from it, the way in which the observer is oriented relative to it, and so forth.” (p 460) but only mentions this in relation to different observers (standing in different relations to the object).

<sup>180</sup> Shoemaker 2002, pp. 462-463.

<sup>181</sup> Ibid., p. 465.

<sup>182</sup> Ibid., p. 466.

Trying to give a pictorial illustration of the OL account of introspection is a bit difficult. The following will have to do:

S has an experience of an apple:  $e \rightarrow \text{apple}$   
S is aware that (of the fact that) she has this experience:  $e \{e \rightarrow \text{apple}\}$

#### 4.5 SUMMARY OF CHAPTER FOUR

Chapter four was intended as an overview of the theories that will be further discussed below. The theories differ from one another in the exact characterisation that is given of qualia. While Shoemaker (and Dretske in one discussion) hold that qualia are representing properties of experiences, Tye claims that they are represented properties of the objects perceived. Lycan is ambivalent and seems to sometimes place qualia in the experience as representing properties, and sometimes with the objects represented. Rosenthal thinks that qualia are intrinsic non-representational properties.

The theories can be divided into two groups; the higher order theories (HO) and the one level theories (OL), respectively. The HO theorists argue that both phenomenal consciousness and introspective awareness should be construed as some kind of higher order representations of the first order experiences. The OL theorists claim that the difference between being phenomenally conscious and not being thus conscious can be accounted for as a difference in the perceptual content. These theorists hold that a proper account of neither phenomenal consciousness nor introspection should entail representations of different orders.

## 5. How many levels do we need?

In this chapter we will disregard the internal differences that exist between the different versions of the HO theories and consider one of their common claims, namely that phenomenal consciousness can be accounted for if we postulate mental states of different orders. The advocates of HO theories claim as great advantages of their theories over one level theories: (i) that a HO theory provides a plausible explanation of the distinction between conscious and non-conscious states by accounting for conscious states in terms of states that the subject is *conscious of* being in, (ii) that the theory can provide an account of the difference between being just aware and being phenomenally aware. We will start by outlining the second claim.

### 5.1 WORLDLY AND EXPERIENTIAL SUBJECTIVITY

A higher order theory of consciousness gives, as its proponents claim, a neat explanation of phenomenal consciousness or the “subjective dimension of experiencing”. Carruthers differentiates between on the one hand *worldly* and on the other *experiential* subjectivity. He claims that the higher order theory can account for the difference between “what the world is like for the organism”, and “what the organism’s experience of the world is like for the organism”.<sup>183</sup> How can we clarify this distinction? Carruthers claims, for a start, that phenomenal consciousness should be equated with *experiential subjectivity*.<sup>184</sup> To possess worldly subjectivity does not make us phenomenally conscious. Carruthers explicates the distinction by using the all time favourite examples of higher order theorists, namely that of the absent-minded truck driver. The truck-driver driving absent-mindedly possesses worldly subjectivity, and when she “comes to” she also possesses experiential subjectivity, Carruthers claims. Say that the driver passes a stop-light which shows green. Her perceptual experience represents the stop-light as green; she will be aware of its green-ness. The driver may drive past this light, on

---

<sup>183</sup> Carruthers 2000a, pp. 127-128.

<sup>184</sup> Ibid., p. 129, note 7.

auto-pilot so to say, thus lacking experiential subjectivity. When the driver becomes experientially aware, she will be aware of what the green stop-light is like for her. It is hence possible to be aware of the green light and not be aware of what it is like to be aware of the green light. An experience of green which it is not like anything for the subject to undergo is said to “possess worldly subjectivity but [to] lack experiential subjectivity.”<sup>185</sup>

Lycan makes a similar distinction. Lycan argues that many philosophers (read: the one level theorists) confuse two different senses of “qualia”. There are according to him one “lower-order” reading of qualia and one “higher order”.

“What, then, of ‘what it is like’? That phrase is now ambiguous, as between phenomenal character, i.e., a quale in the strict sense, and the conscious experience of such a quale, or rather what one knows in virtue of having such an experience.”<sup>186</sup>

Lycan’s distinction between qualia in the strict sense and the conscious experience of qualia fairly well approximates Carruthers’ distinction between worldly and experiential subjectivity, but note that while Carruthers thinks of qualia as narrow contents, Lycan conceives of them as wide contents, at least that is his “official view”. These facts makes Carruthers’ version of the distinction between what states are like in themselves and what they are like for the organism, a bit easier to understand.

The way Carruthers and Lycan use the phrase “what it’s like” may be confusing. According to these thinkers a perceptual experience can be like something even though it is not like anything *for the subject*. A quale “in the strict sense” is like something in itself; so is “worldly subjectivity”. To consciously experience a quale or possess experiential subjectivity means that the experience is like something for the subject. At the heart of the HO theories lies the distinction between, on the one hand, *sensory* or *qualitative* and, on the other, *conscious*. An experience can thus be sensory or possess some qualitative character – it can be like something – even though the subject is not conscious of the experience.

Rosenthal also makes a distinction between two kinds of phenomenal consciousness, between what he calls “thick and thin phenomenality”.

---

<sup>185</sup> Ibid., p. 147.

<sup>186</sup> Lycan 1996, p. 77.

“One kind consists in the subjective occurrence of mental qualities, while the other kind consists just in the occurrence of qualitative character without there also being anything it’s like for one to have that qualitative character. Let’s call the first kind *thick phenomenality* and the second *thin phenomenality*. Thick phenomenality is just thin phenomenality together with there being something it’s like for one to have that thin phenomenality.”<sup>187</sup>

Rosenthal claims that experiences have *sensory qualities*, or *mental qualities*.<sup>188</sup> One of these mental qualities is characteristic for perceptions of green objects. Rosenthal calls this quality “mental green”. A mere “occurrence of qualitative character” is an experience that has a mental quality such as mental green. Now, Rosenthal does not think that mental green represents greenness of physical objects. Mental green is according to him a non-intentional property. If we disregard this disparity between his theory and the theories of Lycan and Carruthers, one can say that the distinction between thin and thick phenomenality corresponds fairly well with that of worldly and experiential subjectivity and with Lycan’s distinction between qualia in the strict sense and conscious experience of qualia (also keeping in mind that qualia are wide contents of experiences according to Lycan and narrow contents in Carruthers view).

Byrne argues that the postulation of higher order representations are unnecessary in order to explain phenomenal consciousness.

“Suppose that one knows – via ‘introspection’, whatever that is exactly – that one’s present experience has property P, and one knows that certain past experiences had property P. Suppose that one knows that yet other past experiences had other properties, Q, R,... Further suppose that experiences possessing these properties thereby saliently resemble each other, and that one knows these facts about similarity. So, for example, one may know that one’s (present) experience is more similar (in this salient respect) to a Q-experience than to an R-experience, and so on. This appears to be sufficient for one to know what one’s (present) experience is like. If it is not sufficient, it is unclear what else needs adding.”<sup>189</sup>

Byrne is thus trying to give a convincing example of what it takes to know what one’s experiences are like. According to him, a plausible account of this

---

<sup>187</sup> Rosenthal 2002b, p. 657.

<sup>188</sup> Rosenthal 1999.

<sup>189</sup> Byrne 2004, p. 220.

can be given without evoking higher order awareness. The crucial term here is however “introspection”. Even though his argument seems very convincing, the absence of qualification of this term undermines it. To know something “via introspection” means, according to the HO theories, to have a higher order representation of a lower order mental state. If this is what Byrne objects to, he needs to specify what it is to know something “via introspection”.

Several philosophers argue that the difference between driving “on auto-pilot” and driving after one has “come to”, need not be described as a difference between, on the one hand, being aware of what the world is like, and, on the other, being aware of what one’s experiences are like. Lycan has e.g., changed his mind about the analysis of this example. Lycan and Ryder<sup>190</sup> claim that the difference between driving on auto-pilot and driving fully consciously can be explained in other ways than in terms of the absence or presence of higher order representations. The difference between the “auto-pilot” driver and the driver who has “come to” can be described in terms of different degrees of attention, they argue. A more alert driver can have a normal level of attention to the road without having higher order representations of her perceptions. It is also possible to drive un-attentively while having higher order representations, they argue. Daniel Dennett argues that the truck driver example can be explained in terms of instantaneous memory loss.<sup>191</sup> Tye claims that the absent-minded driver has phenomenally conscious perceptual states which are *concealed* from the subject.<sup>192</sup>

To the present author, it seems that the fact that the driver has forgotten what has happened a few moments ago could plausibly be explained in terms of diminished attention. It is possibly the case that we forget most things that happen to us during a day, except the things we really pay attention to. When we drive for long periods of times in uninspiring landscapes we tend to fail to pay attention to the environment. The reason why this particular example of the common phenomenon of not-remembering-because-we-have-not-paid-attention is put forward as an argument in these circumstances is probably because the “coming to” is an obtrusive experience in this case. Not paying attention to the traffic can have much more alarming consequences than not paying attention to what one reads, or what someone says or what one sees when one looks out through

---

<sup>190</sup> Lycan & Ryder 2003.

<sup>191</sup> Dennett 1991.

<sup>192</sup> Tye 1995, 1999.

the window normally has. The fact that we do not remember what it was like to drive the car does not conclusively show that we were not phenomenally conscious when driving.

Furthermore, paying attention to the road does not seem to entail paying attention to what the road *is like to us* in any comprehensible sense. Is it not even a fact that to pay attention to what our perception of the road is like draws some of our attention away from the traffic, so that we become less conscious of the road? What would be more effective, driving-wise, paying attention to the cars that one drives past or paying attention to what it is like for me to see these cars?<sup>193</sup>

Another example that Carruthers uses to support the HO thesis is that of “blindsight”.<sup>194</sup> The blindsight example shows, according to Carruthers, that perception can be phenomenally conscious as well as phenomenally unconscious. The blindsighted person has “worldly subjectivity”, but no “experiential subjectivity”, in Carruthers’ view. The blind-sight example is quite different from the case with the truck driver. While the truck driver drives her vehicle with considerable efficiency, the blind-sighted person has very diminished discriminatory abilities. It does not seem accurate to describe this case by claiming that the only thing this patient lacks is awareness of her visual states. She lacks so much more. The only thing being higher orderly aware adds to the equation, according to the HO theorists (at least Lycan), is the dimension of (experiential) subjectivity, it does not add the ability to discriminate and control behaviour.

Carruthers advances a third example to substantiate his distinction between worldly and experiential subjectivity. It is drawn from Milner and

---

<sup>193</sup> Helge Malmgren has pointed out to me that switching from attending to the road to one’s perception of the road may lead to a motion framework shift. Instead of seeing ourselves as moving along the road, we see the road as streaming towards us. To be sure, the latter viewpoint is not optimal for driving.

<sup>194</sup> Patients with damage to the primary visual cortex typically have blind areas in their visual fields. If an experimenter flashes a stimulus in one of those blind areas the patient says that he or she does not see anything. But, some of these patients are able to “guess” certain features of the stimulus having to do with location, motion and direction. (Blindsight was first noted by Pöppel et. al. (1973) and there is now a huge amount of literature on this phenomenon.) In one sense consciousness is missing, while some discriminatory abilities are preserved but in these cases the discriminatory capacities are always degraded compared to those of a normal-sighted person. (See e.g., Farah 1990). The same downgrading of behaviour can also be seen in other awareness-but-no-phenomenal-consciousness candidates like *epileptic automatism* (Penfield 1975) and *prosopagnosia* (Sergent och Poncet 1990).

Goodale's theory of two visual systems.<sup>195</sup> Milner and Goodale's patient, D. F. cannot recognize the orientation of a slot in a disk, whether it is vertical or horizontal but she can post a letter through it. Carruthers claims, in short, that Milner and Goodale's examples of visual dissociation shows that the distinction between worldly and experiential subjectivity is real and that there are conscious as well as unconscious visual experiences (or percepts).

Carruthers, again, argues that the OL theories are inapt to explain the difference between phenomenal consciousness and mere functional awareness. He says that any theory which does not distinguish between "worldly" and "experiential subjectivity" in the way he does, will have the following implications:

"All perceptual states must be conscious ones, on this view, because all perceptual states must have *feel* or must be *like something* to possess."<sup>196</sup>

Does the OL stand entail that all perceptual states have such feel? Perhaps, but the examples Carruthers gives us of perceptual states without feel (blindsight, visual agnosia and truck-driving) do not prove his point either. The absent-minded driver may not be busy attending to her perceptions of her road, but neither does she attend to the road. Her thoughts are elsewhere. And regarding the blind-sight example, phenomenal consciousness is not all the blind-sighted person seems to lack. She lacks a great deal of "access awareness" as well. It is probably the Milner-Goodale example which provides the most powerful of Carruthers' arguments. This phenomenon does not seem to be explained by the role of attention, and it might be the case that a subject has good discriminatory abilities when it comes to behaviour but still no conscious visual experiences. Then again, this is not much empirical evidence to substantiate such a bold hypothesis.

---

<sup>195</sup> Milner & Goodale 1995. Milner and Goodale's patient D.F. is suffering from visual form agnosia which entails that the patient cannot recognize objects or shapes and may be capable of few conscious visual experiences, but the patient's visual sensorimotor abilities are unaffected. Similar dissociations between being visually aware of something and being capable of acting on visual information can also be found in "normal" subjects, which is shown by e.g., the Titchener illusion (illustrated in Milner & Goodale, p. 168). This, among other things, has led Milner and Goodale to postulate the existence of two separate visual systems operating in the brain.

<sup>196</sup> Carruthers 2000a, p. 148.

## 5.2 CONSCIOUS STATES

The higher order theorists hold that the one of the major benefits of their type of theory is that it does not describe consciousness as an *intrinsic* property of mental states. The HO theorists claim that “conscious state” should be analysed as “state we are conscious of”, or “aware of”. Lycan:

“A state of a subject, or an event occurring within the subject, is a conscious state or event, as opposed to an unconscious or subconscious state or event, iff the subject is aware of being in the state or hosting the event.”<sup>197</sup>

Rosenthal says:

“No informative explanation of state consciousness is possible unless we can represent it as having some articulated structure. But it will be hard to justify the idea that being conscious is an intrinsic property of mental states if that property does have some informative structure. Once an explanation assigns such structure, it will be equally plausible to regard being conscious as an extrinsic property of mental states.”<sup>198</sup>

Let us give brief outline of the background for this discussion. The term “conscious” is as we know used in many different but interrelated senses. First, we can use “conscious” as describing an overall state of a person or animal who is awake; we can say that a certain creature is *conscious*, as opposed to being in a coma (or something similar). Secondly, we also say of a certain *state* of a person that it is conscious (in contrast to an unconscious or, perhaps non-conscious state). The first sense picked out here is by some philosophers called “individual consciousness” (Güzeldere 1997), while by others “creature consciousness” (Dretske 1995, Rosenthal 1997). Lycan (1996) names this “control consciousness”, which he contrasts to “organism consciousness” which, in turn, is the *capacity* of being conscious. A human being can be ascribed organism consciousness even when she is in a coma (if there is a possibility that she will wake up again). The second sense, which classifies one’s mental states as being as of one type or another, is labelled “state consciousness” by several philosophers. (E.g.,Güzeldere 1997, Rosenthal 1997, Dretske 1993).

---

<sup>197</sup> Lycan 1996, p. 3.

<sup>198</sup> Rosenthal 1997, p. 736.

“Individual consciousness” can be used both in a *transitive* as well as in an *intransitive* way. People are conscious, period, and they are conscious *of* things. The distinction between *state consciousness* and *creature consciousness* on the one hand and *transitive* and *intransitive creature consciousness* on the other was first introduced by Rosenthal.<sup>199</sup> In order for a person to be transitively conscious it is required that she is intransitively conscious. She can only be conscious of things if she is a conscious being. She cannot be conscious of e.g., the smell of violets and be in a coma at the same time.

Now, the higher order theories hold that a conscious state is a state the subject is *conscious of*. State-consciousness is thus analysed as states the subject is transitively conscious of. That may seem, at least it does to its advocates, as a good theory since it seems to give a straightforward explanation of what it is for a mental state to be a conscious state. The theory has however been criticised in different ways. Dretske formulates one objection to the transitive analysis of consciousness as follows:

“Some people have cancer and they are conscious of having it. Others have it, but are not conscious of having it. Are there, then, two forms of cancer: conscious and unconscious cancer?”<sup>200</sup>

Or as Block puts it: My liver does not become a conscious liver just because I am conscious of it.<sup>201</sup> Being the intentional object of an act of awareness does not seem sufficient for being a conscious object. What is it about a particular kind of state that gives us the right to describe them as conscious just because we are *conscious of* them? Lycan answers that what makes these states special is that they are *mental*.

“Stomachs and freckled patches of skin are not mental. It seems that psychological states are called ‘conscious states’ when we are conscious of them, but nonpsychological things are not.”<sup>202</sup>

Rosenthal claims that states that become conscious when the subject is (seemingly) non-inferentially conscious of them are all states that, in turn, are *about something*.

---

<sup>199</sup> Rosenthal 1986.

<sup>200</sup> Dretske 1995, p. 97.

<sup>201</sup> Block 1994, p. 212.

<sup>202</sup> Lycan 1996, p. 24.

“A state can be conscious only if being in it, even when the state is not conscious, it results in one’s being conscious *of* something, and states of the liver do not qualify.”<sup>203</sup>

This analysis rules out conscious stones and livers but it also seems to rule out conscious moods. Some sentiments do not seem to have intentional objects, like cheerfulness, yet they are conscious. Given Rosenthal’s characterisation of things that can be conscious, such moods cannot be conscious since they, in turn, are not about anything.

How does a one-level theorist like Dretske account for state consciousness? The HO proponents argue that the rejection of a transitive analysis of consciousness entails that consciousness is regarded as an *intrinsic property* of some states. But, as Byrne points out, the one-level theory does not *argue* that consciousness is an intrinsic property of mental states. That consciousness is intrinsic is a plausible consequence of the theory, not the basis of it. He writes:

“To be sure, if you think that intransitive consciousness is an intrinsic (that is, non-relational) property of mental states then of course you will think that the higher-order thought hypothesis is mistaken, for on that account intransitive consciousness *is* relational. Rather, that intransitive consciousness is intrinsic is a plausible *consequence* of the objection, not the basis of it. The objection is that if what makes a state conscious is the fact that it is the object of another mental state, then there is no explanation of why only *mental* states are conscious. For other non-mental states can be the object of mental states, and they are not conscious. This objection plainly does not assume that state consciousness is intrinsic. Rather, it tries to derive an absurdity – e.g., conscious states of the liver – from the premise that intransitive consciousness is relational in the way the higher-order thought hypothesis would have it.”<sup>204</sup>

---

<sup>203</sup> Rosenthal 2002a, p. 722.

<sup>204</sup> Byrne 1997, p. 109.

## 5.2.1 UNCONSCIOUS CONSCIOUS STATES

An implication of rejecting the transitive analysis of conscious state seems to be that there can be conscious states of which the subject is not conscious, a consequence the HO theorists regard as a drawback of the OL theories. Dretske is of a different opinion:

“[A]n experience can be conscious without anyone – including the person having it – being conscious of having it.”<sup>205</sup>

He gives the following example of conscious states of which the subject is not conscious. He presents two constellations of dots, which he calls Alpha and Beta, respectively, and asks the reader to look at them. These figures are dissimilar in a way that Dretske assumes that the reader does not notice. There is one spot in Alpha that is absent in Beta. Dretske calls that spot “Spot”. Now, he lets E(Alpha) stand for one’s experience of Alpha and E(Beta) for the corresponding experience of Beta. Alpha and Beta differ from another, and so must E(Alpha) and E(Beta). E(Alpha) includes E(Spot) as a part which E(Beta) does not. Not only are they numerically different, Dretske claims, they also differ from one another with respect to their intrinsic properties.

Here, the distinction between on the one hand, *object-awareness* and on the other, *fact-awareness* is important. Dretske claims that we are object-aware of the difference between Alpha and Beta but not fact-aware of it. We are hence in a sense *aware of Spot*, but not *aware that* it is present in Alpha but not in Beta. He assumes, thus, that we represent both Alpha and Beta correctly with all their dots but we are not fact-aware of all the dots and therefore not aware that Alpha and Beta have the properties they have.

Now, Dretske claims that this example shows that “state consciousness” and “creature consciousness” (individual consciousness) are independent of one another.

“Once one makes the distinction between state and creature consciousness and embraces the distinction between fact- and thing-awareness, there is no reason to suppose that a person must be able to distinguish (i.e., tell the difference between) his conscious experiences. Qualitative differences in conscious experiences are *state* differences,

---

<sup>205</sup> Dretske 1993, p. 263.

distinguishing these differences, on the other hand, is a fact about the *creature* consciousness of the person in whom these experiences occur.”<sup>206</sup>

Which of these accounts of “conscious state” seems more plausible, the HO or the OL one? To claim that a conscious state can be whatever thing a person is conscious of, does not work as we have seen. Does that entail that we have to think of consciousness as an intrinsic property? That consciousness is an intrinsic property of mental states does not necessarily follow from the rejection of this particular (the higher order) relational account. The property of being conscious can be relational even though it does not express a relation between the state of being conscious and the subject that is conscious of the state. A conscious state can thus be what it is in virtue of its relations to other things. What makes a state conscious, according to Dretske, is the role it plays in making the subject conscious *of some other thing or fact*. Dretske writes:

“An experience of  $x$  is conscious, not because one is aware of the experience, or aware that one is having it, but because, being a certain sort of representation, it makes one aware of the properties (of  $x$ ) and objects ( $x$  itself) of which it is a sensory representation.”<sup>207</sup>

Later, Dretske formulates this account of state-consciousness in terms of acts.

“There are, as far as I can see, only two options for making sense out of state consciousness. Either a state is made conscious by (1) its being an *object* or (2) its being an *act* of creature consciousness. A state of creature  $S$  is an *object* of creature consciousness by  $S$  being conscious of it. A state of  $S$  is an *act* of creature consciousness, on the other hand, not by  $S$  being aware *of* it, but by  $S$  being made aware (so to speak) *with* it – by its occurrence in  $S$  making (i.e., constituting)  $S$ ’s awareness of something (e.g., an external object). When state consciousness is identified with a creature’s *acts* of awareness, the creature need not be aware of these states for them to be conscious. What makes them conscious is not  $S$ ’s awareness of them, but their role in making  $S$  conscious”<sup>208</sup>

---

<sup>206</sup> Ibid., p. 274.

<sup>207</sup> Ibid., p. 280.

<sup>208</sup> Dretske 2000b, p. 183.

Amie Thomasson seems to think that Dretske's qualification of what a "conscious state" actually is, is rather empty. She writes:

"To say that mental states are conscious if they make the creature conscious of the properties (and maybe the object) represented merely pushes the question back a step: What does it mean for a creature to be conscious of something, as opposed to merely having a representation of it? In virtue of what are certain natural representational states, but not others, capable of making their creatures conscious of things in the world? His [Dretske's] only remarks that begin to make such a distinction are that the only natural representations suited to be 'experiences' are those 'whose function it is to supply information to a cognitive system for calibration and use in the control and regulation of behavior' (1995:19)."<sup>209</sup>

One way in which this could be done, continues Thomasson, would be to appeal to the internal phenomenological character of the states. She writes:

"So, I would suggest, there are two senses of 'conscious' at work here: a mental state itself can be (intransitively) conscious in the sense of having a phenomenological character; and that phenomenological character is what can make *us* (transitively) conscious of other things. Ordinarily, the two go together so, e.g., to see *consciously* is to see in such a way that (unlike in cases of blindsight) the seeing has a (visual) phenomenological character, that enables us to be (transitively) conscious of the things that we see. Yet certain states (involving diffuse feelings such as vague depression or an adrenalin rush) may be merely conscious in the first sense without their representing, and making us transitively conscious of, anything else."<sup>210</sup>

This theory thus denies one of the basic claims behind Dretske's theory: that what it is like for someone to perceive something is determined by the intentional content of one's perceptions. According to Thomasson's view, consciousness is intrinsic to mental states and not bestowed upon them by other acts, nor by the content of the states themselves.

I agree with Dretske that what makes conscious experiences conscious may not be best accounted for in terms of the subject's awareness of them, but rather in terms of the role the experiences play in making the subject conscious of the world. But we still need to specify that role. *Why* do these particular states, states with a certain kind of content make us conscious of the world? Dretske appeals to the possible biological function of such states

---

<sup>209</sup> Thomasson 2000, p. 201.

<sup>210</sup> *Ibid.*, p. 204.

in order to provide an answer to this question. He argues that since it is a fact that some states are conscious while others are not, there should be an evolutionary advantage of conscious experiencing.

“The function of sense experience, the reason animals are conscious of objects and their properties is to enable them to do all these things that those who do not have it cannot do.”<sup>211</sup>

This is an interesting line of thought, even though it does not tell us what a conscious state *is*. Investigating what the biological function of consciousness might be, would be a completely different, but perhaps more fruitful strategy.

### 5.3 SUMMARY OF CHAPTER FIVE

One can conclude that it is difficult to explicate exactly what “conscious” in “conscious state” means. The HO suggestion at least seemingly gives a clarification of what “conscious” means, namely “something one is conscious of”. But the explanation also has its costs. As we have seen, being *conscious of* something does not necessarily entail that that which we are conscious of should, in turn, be described as conscious. If I am conscious of a stone or a state of my liver, the stone or the liver-state do not become conscious according to most philosophers including the HO theorists. Only certain kinds of states of conscious creatures exhibit the feature of becoming conscious when the creature is conscious of them and it is difficult to specify what these kinds are. Rosenthal’s story that only states that are *about* something become conscious when the subject is conscious of them does rule out the possibility of conscious livers and conscious stones, but it might also rule out the possibility of conscious moods.

Remember that Rosenthal thinks of qualia or “mental qualities” as non-intentional properties of experiences. Given this view it is conceivable that the subject can be aware of those qualities as his characterisation of “conscious state” demands, but given that qualia are represented properties, it is less clear how the subject can be aware of the state-qua-mental-

---

<sup>211</sup> Dretske 1995, p. 121.

vehicle become conscious of those properties. (We will return to this issue in subsection 6.7, below.)

In this case, accounting for conscious states in terms of what they *do* for the system instead of what they *are*, and specifically, what the biological function of *creature consciousness* is, is perhaps a better strategy. Is there some competitive advantage of being phenomenally aware of the world? Persons who suffer from blind-sight seem at all events slightly disadvantaged and driving un-attentively may also be a bit disadvantageous...

## 6. Against HOP

Below I will review some arguments which have been influential in these discussions of introspection and representationalism, as well as present some of my own. First we will consider possible arguments against HOP.

A vast number of arguments against the claim that introspection is an inner sense have been advanced by different philosophers and at different times. Below, I will limit the discussion to those arguments which in my view are the most telling, and overlook others.<sup>212</sup>

### 6.1 THE ALLEGED INTROSPECTIVE ORGAN

One recurrent argument against the higher order perception theory is based on the apparent absence of an introspective sensory organ. If introspection were anything like perception, the story goes, it would entail the presence of an introspective sense organ. Just as we see with our eyes and hear with our ears, and feel with the skin, there must also be something we introspect *with*.

The first counter question that comes to mind is: are we absolutely certain that there is no such organ? The fact that we haven't found one yet does not exclude the possibility that there is one. We would perhaps have a stronger argument at hand, if it could be maintained that an introspective

---

<sup>212</sup> Rey (1983) argues that inner sense cannot be sufficient for phenomenal consciousness since computers can deploy higher-order representations of their first order states while we do not choose to call them conscious. Lyons (1986) and McGinn (1996) argue that small children lack introspective abilities, but are still phenomenally conscious. If phenomenal consciousness is dependent on some ability of inner perception, children would learn how to introspect at the same age they learn how to perceive. Carruthers (2000a) argues that the HOP theory is hard to defend given the accuracy of the evolutionary theory. The internal monitors postulated by these theories would require considerable computational complexity in order to do their job. The organism would need to have mechanisms to generate a set of internal representations representing the content of the experience. The scanner, which supposedly generates higher-order experiences, would have to be just as sophisticated and complex as the perceptual systems themselves are, Carruthers claims, and it is difficult to find a story that explains how the evolutionary pressures could ever had led to the construction of such a system. Dennett (1991) argues that the inner sense model presupposes a "Cartesian theatre", a place in the brain where it all comes together and where "it" can be observed by a physically realised Cartesian soul.

organ is *impossible*, that it does not exist in any possible world. Auguste Comte writes that the existence of an introspective organ is precluded, since no organ *can observe itself*.<sup>213</sup> I don't find this convincing, what reasons do we have for claiming that a self-observing organ is impossible? In a trivial sense of self-observation, an eye can watch itself through a mirror. In a slightly less trivial sense, we can imagine an organ whose very task is to monitor its own activity. I am thinking about something that resembles an eye, but where the incoming stimuli are the previous output. I imagine a nerve which takes the output from "the retina" as its input and then feeds the "eye" with the same information once more.

Maybe Comte would not consider my organ as a self-observing organ in the first place. Quite possibly, the organ that he thinks is impossible is an organ which performs a certain non-trivial function and monitors these operations at the very same time that it is performing them. But nor is such an organ impossible. It is an empirical fact that a human being is limited when it comes to performing certain tasks simultaneously; it is e.g., very difficult to perform head calculations while writing a poem at the same time or read a text out loud while writing another. Maybe we are incapable of entertaining a thought and observe it at same time due to the same factors, that we have limited attentional resources or something similar, but not because it is impossible in principle (if that is was Comte had in mind) to introspectively observe oneself.

Then again, why does the alleged introspective organ have to observe *itself* in the first place? Is it because the observer and the observed one are part of the same mind? Or it is because the brain is one single organ? If this is the reason, it would be equally impossible to feel one hand with the fingertips of the other, because the surface of the hand and the finger are part of the same organ - the skin. Materialists who argue that we have an introspective "organ" do conceive of one part of the brain examining some other part.

If we put the emphasis on "sense", it can be argued that an introspective sense organ is impossible, since it lies in the definition of a sensory organ that such an organ is sensitive to and hence registers information about *things outside the nervous system*. By characterising sensory organs thus, we have effectively ruled out the possible existence of an introspective organ since

---

<sup>213</sup> Comte 1830. I am not sure what kind of impossibility Comte is referring to, whether an introspective organ is logically impossible or nomologically impossible (or perhaps metaphysically impossible).

the task of this organ is to register intra-neural or intra-cognitive information.

Could we perhaps give another definition of “sensory organ” which would not exclude the possibility of an introspective organ in this way? In biological terms an organ is a structure in an animal which performs or has the capacity to perform a task which plays an important role for the survival of the animal, or at least for the well being of the host (it has a function). An organ has a specific task, which is to the benefit of the animal, like registering information from the outside world, or to purify the blood. In order to describe an organ it seems as if we first have to decide what the function of the organ in question is. The function of an introspective organ would probably be to register information about what is going on in the mind in order to contribute to the overall functioning of the person. We cannot for a priori reasons rule out the existence of such an organ. But, then again, the existence of an organ in this (non-sensory) sense does not entail any answer to the question whether or not introspection shares any important features with *perception*. Given this definition of organ, the possible existence of such an introspective organ rather suggests that introspection may as well share some interesting similarities with e.g., the workings of the cardio-vascular system.<sup>214</sup>

Armstrong faces up to the organ-related critique by simply claiming that the requirement of an introspective organ is uncalled for. There is no reason why we should compare introspection to *vision*. The most accurate analogue for introspection is “bodily perception” (proprioception – the sense of the position of the body in space), according to Armstrong. Higher order perception resembles bodily perception in two respects; in both cases there is nothing we perceive *with*, and secondly, higher order perception just like bodily perception is “private”. The object – our own body – is private to each perceiver.<sup>215</sup> This privacy is in both cases an entirely contingent matter,

---

<sup>214</sup> Another argument against the HOP, put forward by e.g., Carruthers 2004 and Sturgeon 2000, has to do with sensory organs’ proneness for malfunctioning. Eyes and ears, e.g., are not altogether reliable with regards to how they register and represent information to the subject. If the faculty of introspection uses an organ of this kind, introspective knowledge cannot be *infallible*. As we have seen some philosophers gladly accept this implication, the HOP proponents being some of them. Critics, on the other hand, argue that since introspective knowledge is infallible, it is at least highly unlikely the introspective “function” is performed by an organ.

<sup>215</sup> Armstrong 1968, p. 325.

he adds, it is neither metaphysically nor conceptually impossible for someone else to perceive the happenings in my body or in my mind.

Visual perception appears to constitute the standard example of perception. I think it is fair to say that when no further characterisations are given in a text, the reader can be expected to assume that perception is visual perception. Therefore it is quite likely that when introspection is compared to perception, it is visual perception that is being thought of. There are, after all, few descriptions of the “inner nose” or the “mind’s ear” in the literature. The interpretation of perception as visual perception is understandable since vision is the epistemically most important sense modality for sighted persons.<sup>216</sup>

Given these facts as well as how the HOP theory is usually phrased, it seems a priori probable that the HOP theorists have vision in mind when they compare introspection to perception. Armstrong, however, explicitly denies this. But when we look closely at the accounts of inner awareness given by both Armstrong and Lycan, it seems as if bodily perception lacks one important feature that the inner sense supposedly has. According to the HOP theory a first order state can exist while not being attended to or perceived. Proprioception, on the other hand, cannot be “shut off” in the way that sight can be when we close our eyes. These features of proprioception and sight respectively, gives us good reason to think that the latter sense modality provides a better analogy for the inner sense. The chief claim of all the higher order theories, including Armstrong’s, is that the presence or absence of inner awareness makes a difference to consciousness. Therefore it could be argued that the “external” sense that Armstrong mentions as the correct comparison to inner perception does not have the desired features to work as an analogy for inner sense.

To conclude;

(i) If we define “introspective organ” as a sensory organ which observes the very same operations it performs, it might be precluded by conceptual necessity, since sensory organs by definition do not monitor themselves. But that is not what the proponents of the HOP theory imagine, they claim that

---

<sup>216</sup> C.f. Bruce & Green 1996. (Maybe we should, after all, be surprised that there are no allusions to the mind’s ear in the philosophical literature. Do we not e.g., “listen” to inner music and, when having conversations with ourselves, do we not also “listen” to what we have to say to ourselves?)

the alleged organ is one part of the brain that scans another. Such a definition is hence uncalled for.

(ii) If we claim that the information a sensory organ processes must come from “outside” (either in spatial or in metaphorical terms) the nervous system, an introspective organ is precluded, but the presence or absence of an organ in that respect does not seem to be relevant to the question of introspection being an *inner* sense or not.

(iii) If it is sufficient for a definition of “introspective organ” that it has a function to perform on behalf of the subject’s well-being, the question of such an organ has nothing to do with introspection being a kind of inner perception or not.

(iv) Not even all our “ordinary” perception requires the use of an organ. The proprioceptive sense (our sense of the relative location of our limbs) is an example of a sense modality without an organ.

The unlikelihood of the existence of an inner sense organ has, thus, little bearing on the issue whether introspection is an inner sense or not. And, it should be added: few philosophers maintain that the improbability of there being an introspective sense organ in itself provides a knockdown argument against HOP.

## 6.2 SPECIAL PHENOMENOLOGY

It is tempting to argue that another important feature of perception is that perceiving, unlike e.g., thinking, entails some special qualitative character. It is like something to perceive while it is not like anything to think about something. It is also often maintained that the different sense modalities differ from one another, partly in virtue of how it feels to perceive in them. It feels differently to see a tomato than to taste it. Introspective experiences, on the other hand, do not distinguish themselves by featuring some unique

phenomenal character. It does not feel differently to look at a tomato and to introspect this perceptual experience.<sup>217</sup>

It can however also be argued that some special phenomenal character is not necessary for something to count as a special sense. According to a tradition that goes back to Aristotle, human beings have five senses: taste, smell, touch, sight and hearing. Potential new sense modalities are proprioception, the vestibular sense (the sense of one's orientation with respect to gravity) and also perhaps pain. (According to some scientists, they may even be more. Robert Rivlin and Karen Gravelle claim that a more accurate count would probably give seventeen senses!)<sup>218</sup> The idea that special phenomenology (or a special set of qualia) and sense modality go hand in hand may seem as a plausible suggestion as long as we stick to the traditional five senses, but if we want to extend the number of senses to at least seven (including the proprioceptive and vestibular senses) it seems doubtful that special phenomenological character could be considered a necessary criterion. It does not feel in any specific way e.g., to perceive the position of your limbs. On the other hand, it is not beyond dispute that proprioception completely lacks phenomenology. The special character of proprioceptive perceptions may be less obtrusive than that of pain perception or visual perception, but it may still be there. One might also argue that introspection *does* have a special phenomenology however faint, perhaps like proprioception. It could be the case that there are introspective feelings that are so vague and indistinct that we don't notice them unless we are really good introspective observers. (To find them we would have to introspect the products of introspection, which would be a third order kind of consciousness.) There might for instance be a feeling of mine-ness and/or one of effort as described by William James and more recently by e.g., Graham & Stephens (1994) and Mangan (2001). But if these feelings exist they are very mild indeed. When I turn from attending to the white horse to my impression of the white horse the predominate feature of that act is still the horse.

Both Lycan and Armstrong maintain that there is no reason why we should expect an introspective phenomenology at its own level of operation.

---

<sup>217</sup> This point has been put forward by e.g., Lyons (1986), Rosenthal (1997, 2002), Dretske (2000) and Güzeldere (1997).

<sup>218</sup> Rivlin & Gravelle 1984.

Lycan writes:

“[W]e should not expect internal monitoring to share the property of involving some presented sensory quality at its own level of operation. For the sensory properties presented in first-order states are, according to me [...] the represented features of physical objects; e.g., the color presented in a (first-order) visual perception is the represented color of a physical object. First-order states themselves do not have ecologically significant features of the sort physical objects do, and so we would not expect internal representations of first-order states to have sensory qualities representing or otherwise corresponding to such features.”<sup>219</sup>

Lycan emphasises that he does not mean that inner sense is exactly like any of the outer senses. Giving an account of what it is to perceive that is tailored to fit the classical five senses and nothing else, and then showing that the inner sense does not comply with this account would hence not constitute a good argument against the HOP theory. Quite possibly, it has never been anyone’s intention, (not Lycan’s nor Locke’s, e.g.,) to claim that introspection or inner sense should be regarded as “a sixth sense” or “an eighteenth sense”, but rather as “*the* inner sense”, compared to “outer sense”. Still there must be some features of this inner awareness that motivates the analogy with perception.

### 6.3 IMMUNITY TO ERROR

One feature which higher order perception presumably shares with outer perception is that is fallible. According to the HOP thesis the subject is not always accurate in her beliefs about her own experiences. It is hence possible to mistake one (kind of) experience for another (kind of) experience. The subject may e.g., believe that she perceives a red tomato, when she in facts perceives a green tomato. This claim has been criticised by several philosophers and by Shoemaker in particular. He argues that can we cannot be mistaken either about the modes or the contents of our mental states. His argument partly builds on the assumption that there is no room for misidentification of a mental state, and therefore neither of identification. In comparison, sensory perception gives the observer the possibility to

---

<sup>219</sup> Lycan 1996, pp. 28-29.

distinguish one object from others and to re-identify the same object over time.

### 6.3.1 Misidentifying the content and attitude of a mental state

Shoemaker writes:

“I am aware that I believe that Boris Yeltsin is President of Russia. It seems clear that it would be utterly wrong to characterize this awareness by saying that at some point I became aware of an entity and identified it, that entity, as a belief that Boris Yeltsin holds that office. To say that would suggest that it ought to be possible for someone to become aware of a belief and misidentify it as something other than a belief, or as a belief with a content other than the one it has.”<sup>220</sup>

The reason why this is not possible is that intentional states such as beliefs do not, according to Shoemaker, have intrinsic, non-relational properties, or properties of any kind by means of which they could be identified.

This argument has been criticised by e.g., Cynthia MacDonald.<sup>221</sup> She claims that it *is* possible to make mistakes with regards to the *contents* of our mental states, in the same way that we can be mistaken about what physical object we are acquainted with in perception. As an example of this she presents a case where a person mistakes one “memory” for another. It is conceivable, she argues, that a person can think that she remembers one thing when, in fact, she remembers quite another. She presents the following example: a person takes a trip to Oxford and looks at a particular building there. At the time for her trip she does not mistake the building for something else. She is not near-sighted, the weather conditions are perfect for visual perception and she can, and does as a matter of fact, identify the building she sees as a college house. A couple of years later when she looks back upon the visit to Oxford, she thinks she saw a church. This, MacDonald claims, could be an example of misidentifying the content of a recollection state.

This example assumes a lot. Since it should be a case of mis-introspection and not of mis-remembering, we must assume that the person mistakes an occurrent mental state with the content [college building] for an occurrent state with the content [church]. Secondly, given that we can introspect occurrent memory states in this manner, the actual content of the

---

<sup>220</sup> Shoemaker 1996c, p. 213.

<sup>221</sup> MacDonald 1999.

occurrent memory state must be [college building] and the mistake must arise when the person is introspecting this memory. Given that the content of memory states is wide, it seems improbable that a memory could be either identified or misidentified, just as Shoemaker argues.

Furthermore, MacDonald's example entails that the person did not misperceive the building, but she does not explicitly rule out the possibility of misremembering. It seems probable that this incident could also be explained in any of the following ways:

(i) The person did not properly "store" the perceptual experience in the first place, and therefore the specific mental state that she is now entertaining is not a memory of the building.

(ii) The stored memory got lost on the way, it might have been erased somewhere down the line, and again, the occurrent state by means of which she seems to remember *the building* is not a stored representation of the actual building.

(iii) She was unable to retrieve the right memory in the remembering process. By mistake she got hold of another one.

In order for the situation to exemplify mis-identification by introspection, we must first rule out all these possibilities, which MacDonald doesn't. Furthermore, all three suggestions seem much more plausible than the explanation MacDonald gives.

It is also possible, MacDonald continues, to misidentify one *kind* or mode of mental state for another. She writes: "one might misremember an experience of anguish as one of embarrassment."<sup>222</sup> Again, it is not ruled out that it is not an error of recollection. Is it also possible to be in anguish and believe that one is embarrassed? I will leave that question open here.

### 6.3.2 Misidentifying the self

Shoemaker claims that it is most important for the inner perception theorist to show that we can introspectively identify *the self*, since experiences, he says, are *adjectival* on a subject. The ontological status of an experience resembles that of the bending of a branch or the rising of the sun. He writes:

---

<sup>222</sup> Ibid., p. 720.

“It hardly makes sense to suppose that there could be a mode of perception that has as its objects bendings of branches or risings of the sun, but never branches or the sun. And it makes equally little sense to suppose that there might be a mode of perception that had as its objects experiencings but never experiencers.”<sup>223</sup>

And he adds further:

“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.”<sup>224</sup>

Shoemaker argues that in order for successful identification of any object to take place, the possibility of error must be provided for. There are some cases when it is not possible for the observer/thinker to gain information which would either sustain or contradict her identification of an observation/thought. One of these cases involves statements using the first-person pronoun, as e.g., “I am in pain”. That sentence and those akin to it are *immune to error through misidentification relative to the first-person pronoun*, a phrase coined by Shoemaker.<sup>225</sup>

---

<sup>223</sup> Shoemaker 1996b, p. 10.

<sup>224</sup> Ibid.

<sup>225</sup> Shoemaker’s discussion of immunity to error through misidentification is embedded in a field of related problems concerning the intricate features of the self and of referring to a self. In 1968 Shoemaker published his paper “Self-reference and self-awareness” which together with Henri-Hector Castañeda’s paper “‘He’: A study of the logic of self-consciousness” from 1966 formed the starting-point for a modern discussion of the difference between awareness of the self and awareness of other things, and how the use of the first person pronoun differs semantically from the use of third person pronouns. Frege (1977 [1918-1919]) Wittgenstein (1953), and Anscombe (1981) have previously discussed these topics. We can sort out two different, however intimately related, issues in this area. They are:

(i) In certain situations we are immune to error through misidentification of another as oneself.

(ii) When we refer to ourselves using “I”, this often takes place without us identifying any thing as ourselves via properties that we have ascribed to that thing.

At first glance it may seem as if the term “I” works just like any singularly referring term. One thing that the proposition which is expressed by the sentence “I am in pain” seems to have in common with propositions that are expressed by: “He is in pain” or “Bertil is in pain” is that they all contradict “No one is in pain” and sustain “Someone is in pain”. I.e. “I am in pain” behaves logically as a value of the function “x is in pain”. Still, several philosophers have come to the conclusion that “I” is not a referring expression at all. Frege claims that it is a referring

Shoemaker's point is that since we seem to be unable to misidentify ourselves in this way, we can also not be said to be able to identify ourselves, and in so far as we do refer to ourselves we do it in a unusual way, namely without identifying ourselves. Wittgenstein writes:

“To ask ‘are you sure that it’s you who have pains?’ would be nonsensical. [...] And now this way of stating our idea suggests itself: that it is impossible that in making the statement ‘I have a toothache’ I should have mistaken another person for myself.”<sup>226</sup>

It is the “I” in the sentence “I am in pain” which is the concern here, not that being in pain is something that cannot be verified. Wittgenstein's main point here is not that these statements are *totally* immune to error but only that they are immune to error of a certain kind, namely those that are due to misidentification of a *person*. Consider the case where I say: “I have a toothache”. Even though I might mistake the *sensation* I have for some other (kind of) sensation, as well as I might be mistaken about the *location of the sensation*, maybe it is not in my tooth, but in my jaw, I cannot mistake myself for some other person. We are immune to error in this regard, not because we are so good at judging who we are, but because we are not in a position where we could get it either wrong or right.

Peter Strawson and Peter Geach both argue that in *solitary* circumstances, the term “I” is superfluous. Even though I cannot mistake myself for some other person, the word “I” still identifies to *other persons* the subject to whom the predicate applies. The question: “is it you who are in

---

expression but with a very special and primitive meaning. Anscombe argues that the term does not refer at all, it is a “grammatical placeholder”, like “it” in “it rains”. Gareth Evans (1982) entertains the same basic idea as Anscombe. He remarks that we cannot properly refer to the I since we do not know what the I is. In order to successfully be able to refer to an object, the referrer has to be able to distinguish this object from every other object. Evans explicates this principle thus: a necessary condition for identifying something (and hence for failing to do so) is the possibility to know that something is *F*, but falsely believe that it is *a* that is *F*, when it is really *b* that is *F*. Real identification must involve at least two propositions. It is hence only when a person's knowledge that *a* is *F* can be deduced from knowledge that *b* is *F* and that  $a=b$  that we can speak of a successful identification of *a*. Basically we need to be able to separate the holder of properties from the properties it holds. If we don't have any means for varying subject and predicate, substance and properties, we have not been able to properly identify any substance.

<sup>226</sup> Wittgenstein 1969 [1933-4], p. 67.

pain?” is then informative, because it could be *someone else* who is in pain. This question could e.g., be put by a doctor entering the examining room where two people are waiting. Geach claims that the word “I” is a communicative device, and not something we can use in thinking about ourselves.<sup>227</sup> Strawson makes a similar suggestion, he says that it is correct to speak of self-ascription, for instance of pain, to tell others that one is in pain. If one, on the other hand, were alone, one’s judgment could equally well be of the form: ”there is a pain.”<sup>228</sup>

### 6.3.2.1 Experienced self-identification

Shoemaker argues that we cannot identify our own selves, which is shown by the fact that we cannot make any sense of first person present tense reports where the subject mistakes herself for some other. But it is not only in language that this immunity is revealed. It is also a *phenomenological fact*, Shoemaker claims, that we do not find a self when we try to “look inwards”. David Hume writes:

“[W]hen 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 and pleasure. I can never catch myself at any time without a perception, and can never observe anything but the perception.”<sup>229</sup>

Hume did not find a self by looking inwards. But here I would like to ask: how did he know what to look for? If he did not have an idea of what it would be like to encounter a “self” prior to the searching, how did he know that he did not see it? Hume admits that he has no idea of “self” and so, presumably, no idea of what it would be like to introspect one. He argues that he has no idea of a self because he has not “seen” the self. Is it not also possible that since he did not know what to look for he missed it? Shoemaker’s argument seems to assume, just as Hume possibly did, that there should be some special phenomenological features associated with introspective identification of the self. If Armstrong and Lycan are right, inner perception must not necessarily be accompanied by special phenomenology, in fact it does not have to reveal any phenomenological data at all. If we, once again, assume that higher order perception resembles

---

<sup>227</sup> Geach, 1957, ch 26.

<sup>228</sup> Strawson 1959, pp. 99-100.

<sup>229</sup> Hume 1978, [1711-1776] p. 252.

proprioception, we could argue that identifying the self is like identifying our own foot in proprioceptive space. The perceptual awareness of the location of the foot is not phenomenal. Yet, we do not infer that we cannot identify our foot just because it is *nothing it is like* to be aware of its location.

### 6.3.2.2 The self and the world

Given that the self *is* immune to (perceptual) misidentification, does this really imply that we cannot observe “the self”? To begin with, the conclusion rests on the assumption that the self’s counterpart in the domain of ordinary perception is *not* immune to error, and naturally, this is Shoemaker’s view. Remember that Shoemaker claimed that the relation between mental states and the self is analogous to that of properties and the objects of which they are properties. Natika Newton claims that the analogy Shoemaker assumes between perception and introspection is inaccurate. The fact that I can be uncertain about which physical object I am perceiving in a way in which I cannot be uncertain about which self I am introspecting does not exclude a perceptual account of introspection, since the objects at issue do not play parallel roles in the two modes. She writes:

“It is not that introspection is about the self in the way that perception is of many objects. Rather, *each* mode presents a manifold of elements within a domain, and the elements of each domain, and the domains themselves, are radically different. In the mode of perception the domain is the external world or external reality, and the elements are physical objects. In the mode of introspection the domain is the self, and the elements are states of the self.”<sup>230</sup>

On this way of conceiving introspection, it is “self” and “external reality” which are parallel rather than “self” and “physical object” and the immunity to error through misidentification now applies equally well to both domains, Newton argues. The fact that I cannot mistake myself (as a subject) for another self does not show that introspection cannot be quasi-perceptual, because neither can I mistake the external world for another world. The reason that there seems to be immunity in introspection and not in perception, Newton says, is that in the former case we are looking for errors *across* domains, while in the latter *within* a domain. In fact, the possibility of

---

<sup>230</sup> Newton 1988, pp. 26.

inter-domain error and the impossibility of intra-domain error are similar in the two cases.

“External objects are just as immune to being mistaken for states of myself as my mental states are to being mistaken for objects in external reality. And I, as the domain of my inner experiences, am no more likely to mistake myself for some other domain (some other I) when I am paying attention to these experiences, than external reality is likely to be mistaken for some other reality.”<sup>231</sup>

This is an interesting objection. It could be argued, in Shoemaker’s defence, that external *objects* are not as immune to being mistaken for “states of myself” as my mental states are to being mistaken for objects in external reality, (or someone else’s mental states). We might e.g., question the reality of external objects in a meaningful way, wondering whether they might be imaginary. Assume that I look at an empty street, turn my head away for just a second and then turn back to look at the street again. And, suddenly there is a car standing there. I feel sure it wasn’t there a moment ago and start to question my senses and even think that I am just imagining a car standing there. If the car, in fact, is a real car, I am mistaking the physical car for a “mental car”. In a sense I have mistaken the domain to which the object belongs. The question is: have we mistaken the external *world* for some other world? I do not have an answer to this question.

In sum, it might be possible that the self is immune to misidentification in a way that external objects are not, which could entail that we do not know ourselves qua object of experience. Introspection, thus, would lack one of the important features of perception, namely that of yielding identification information about the object. One might ask, though, whether the ontological status of experiences is comparable to that of bendings or risings in that they cannot exist (for conceptual reasons) without selves, branches and suns? And one might object, as Newton does, to the comparison between the self and physical objects. If the self is on a par with the external world, the same immunity *may* prevail in both domains.

---

<sup>231</sup> Ibid., pp. 26-27.

### 6.3.3 Misidentifying the phenomenal character of an experience

Lycan and Armstrong do not only claim that the subject may mis-introspect the content and the attitude of a particular experience, they also argue that it is possible to be mistaken about the qualitative character of a perceptual experience. Lycan exemplifies this with the razor blade example. (See subsection 2.2.1.)

To other philosophers, the possibility of being mistaken with regard to the nature of our perceptual experiences implies an absurd distinction between the *appearance* of an appearance and the appearance *itself*.<sup>232</sup> Shoemaker writes:

“No one thinks that in being aware of a sensation or sensory experience, one has yet another sensation or experience that is ‘of’ the first one, and constitutes its appearing to one in a particular way. No one thinks that one is aware of beliefs and thoughts by having sensations or quasi-sense-experiences of them. And no one thinks that there is such a thing as introspective sense-experience of oneself”<sup>233</sup>

Is that so? Is there *no one* who thinks that? What *do* Armstrong and Lycan think then? Lycan writes:

“[T]he inner-sense theory implies an appearance/reality distinction for subjectivity.”<sup>234</sup>

And:

“I agree that the notion of an appearance/reality distinction for conscious awareness is odd on its face, and I am inclined to think that dramatic cases of an appearance/reality gap are rare and pathological, but I see here no powerful objection to the inner-sense view.”<sup>235</sup>

So, it seems as if this is exactly what Lycan thinks. “You might introspect a sharp, severe pain when there is in fact no pain at all”, he writes.<sup>236</sup> Shoemaker possibly means that even though Lycan thinks that he thinks that this is possible he cannot truly believe this, since an appearance/reality

---

<sup>232</sup> E.g., Byrne 1997, Harman 1990 and Lyons 1991.

<sup>233</sup> Shoemaker 1996c, p. 207.

<sup>234</sup> Lycan 1996, p. 17.

<sup>235</sup> Ibid., p. 23.

<sup>236</sup> Ibid., p. 19.

distinction for sensations is, in fact, conceptually impossible. Since Lycan and Shoemaker differ in their opinions about what a sensation *is*, what Lycan holds may not be quite as impossible as what Shoemaker thinks that he holds. Shoemaker argues that to feel pain and at the same time not feel pain is impossible. Lycan, on the other hand, thinks that a pain can be unfelt but still be a pain with all its properties. We will come back to this disagreement right below.

## 6.4 IMMUNITY TO IGNORANCE

We are not aware of everything that goes on in front of our eyes. Some events pass by us unnoticed because we are simply not paying attention to them, but there are also features of the world that cannot be perceived at all by creatures such as us. Some physical objects are too small or too fast to be apprehended by human beings. There are also energy forms to which we are not at all sensitive, and any information those may transfer we are not able to pick up. Does introspection resemble perception in this respect? It is certainly a fact that there are many processes and states inside *our bodies*, which we are unable to become *introspectively* aware of, as e.g., cell metabolism or synapse activity. If the object of introspection is considered to be the states of the body, we would have an analogy here. Introspection just as perception could be described as sensitive to only a selection of physical information. And we can imagine other living creatures that could introspect their cell metabolism and liver functions, for example. If we, on the other hand, define the possible object of introspection as “the inner world” in a Cartesian sense, being unable to introspect the functions of liver is not analogous to our inability to sense e.g., electro magnetic fields, since the inner soul is partly *defined* by our introspective access to it. Our inability to perceive electro magnetic fields is however not excluded by definition, since the physical world is not defined with regards to what we can detect by means of perception.

Shoemaker argues that it is impossible to be “blind” to our sensations. While one can be blind to trees, mountains, horses, etc. one cannot be introspectively blind to one’s own experiences of e.g., trees or to one’s conscious beliefs in the existence of trees. Shoemaker invites us to imagine creatures that have intellectual powers comparable with ours and who also have pains like ours but are introspectively blind to those pains. The only

access they have to these pains is from a third person point of view. They can observe their own pain behaviour from “the outside”. Now, Shoemaker also imagines that their pains are unpleasant, even if the creatures don’t *feel* them. Pain would still play the same functional role, the self-blind creature who is in pain would dislike its pains and wish for them to go away, it is just that it is not aware of the actual (feeling of) disliking. The disliking would also have the normal behavioural consequences. The creature reaches for a painkiller or dials the number to the doctor when it is in pain.

The idea of pains that does not hurt anyone is inconceivable, Shoemaker says. “To such a creature”, he continues, “it should seem as if someone else had taken possession of its body.”<sup>237</sup> The relation between pain and awareness of pain is therefore necessary, Shoemaker argues. A pain that does not really hurt anyone is inconceivable and pain avoidance behaviour is only intelligible in the presence of pain awareness.<sup>238</sup>

“Indeed, to say that a creature wants to be rid of pain presupposes that it believes that it is in pain. One can want not to have something while being agnostic about whether one has it; but one can’t want to be rid of something, to *cease* having it, without believing that one currently does have it.”<sup>239</sup>

Lycan rejects Shoemaker’s conclusion. He claims that Shoemaker is “running together two senses of ‘pain’”, namely the first-order pain condition and the second-order *awareness* of this condition. He writes:

“Of course, it is impossible for any creature (ever) to have a pain in the strong, composite sense and yet remain unaware of it. But Shoemaker seems to admit that our pain-blind creatures can have pains in the weaker, first-order sense; he just chooses to withhold the word.”<sup>240</sup>

Remember that Lycan distinguishes between, on the one hand, qualia in “a strict sense” and, on the other, conscious experiences of qualia. It seems as if Shoemaker argues that “strict pains” are not pains in a psychological sense,

---

<sup>237</sup> Shoemaker 1996c p. 227.

<sup>238</sup> Shoemaker admits (1996: 227) that it might happen that e.g., an injured athlete does not notice his pain until he is put on the bench, but he still seems to think that he can maintain that the connection between sensations and the subject’s awareness of those sensations is one of conceptual necessity.

<sup>239</sup> Shoemaker 1996c, p. 228.

<sup>240</sup> Lycan 1996, p. 19.

they would just be states of tissue damage. The HOP model entails that second order awareness is awareness of first order *mental* states. If Shoemaker is right, the theory of inner sense is really a theory of outer sense. If we have a certain physiological phenomenon which we call “pain”, that does not give rise to any behaviour on the subject’s behalf and does not feel in any way to the subject, there is really no reason why we should call this pain a *mental* state or episode.<sup>241</sup>

A main assumption behind Shoemaker’s arguments for immunity to error and ignorance, respectively, is that our mental states are not independent of our introspective awareness of them. Shoemaker describes two “models” of perception: the “object perception model” and the “broad perceptual model”. The object perception model supposes that knowledge of facts about external things is mediated by direct awareness *of* external objects. The broad perceptual model only claims that the objects and states of affairs which are perceived exist independently of us perceiving them. It is not the case, he says that everything we call “perception” conforms to the object perceptual model.

“The sense of smell, for example, does not ordinarily put one in an epistemic relation to particular objects about which it gives identification information.”<sup>242</sup>

Olfactory perception seems still to fulfil two conditions for perception: Our perceptual beliefs are caused by states of affairs that are perceived, and the objects and states of affairs one perceives are taken to exist independently of

---

<sup>241</sup> In “Self-Knowledge and Self-Identity” (1963, Ch 3, Sect 5), Shoemaker puts forward a slightly different version of his “independency objection”. He points out that it is a contingent matter whether a person perceives a certain object or not. This implies that I can perceive Jones, perceive a tree and perceive whether or not Jones perceives the tree. In introspection it ought thus to be possible for me to perceive myself, perceive a particular mental state I am in and perceive whether or not I am aware of that mental state. But it is impossible that I should perceive that I was not perceiving that state, because then I must be perceiving it, he argues. Armstrong replies that if I would come over a piece of unwelcome information and scrutinize whether I have fully accepted this, it would be possible that I come to realize that a part of my mind has rejected it. “Part of my mind ‘perceives’ that another part of my mind fails to perceive the truth of certain information.” (1968, p. 332.) Shoemaker may object, which Armstrong acknowledges, that since part of my mind perceives the truth, I perceive the truth. But Armstrong insists that it is possible that I perceive the truth and also do not perceive the truth, and perceive that I don’t perceive the truth.

<sup>242</sup> Shoemaker 1996c, pp. 222-223.

there being creatures with the capacity to perceive them. The argument from immunity to error is built on the implausibility of introspection conforming to the object perception model, while the argument from immunity to ignorance intends to show that neither does a proper account of introspection meet the requirements of perception in the broad sense.

## 6.5 DISTORTION OF THE OBJECT

Christopher Hill argues that an important disanalogy between perception and introspection is that introspecting an experience may change the experience, while perceiving an object does not change the physical object.

“[E]xtramental entities can exist without standing in any informational relations to the physical eye. [...] The inner eye hypothesis claims that the same things are true, *mutatis mutandis*, of sensations and one’s internal scanning device. It asserts that sensations can exist without being scanned, and also that the internal qualities of sensations do not change when one scans them.”<sup>243</sup>

Hill claims that sensations often change properties because we attend to them. He provides two examples of this: “volume adjustment”, and “activation”, respectively. Hill argues that attention may change the phenomenal volume of an auditory sensation, the severity of a pain, or the strength of a feeling of pressure. One can also activate a sensation, he claims, instead of merely modifying one.

“Activation occurs if one succeeds in actualizing [...] a sensation of the right sort. Thus, for example, having lost touch with the aftertaste of one’s most recent cup of coffee [...] one might suddenly recall the aftertaste and undertake to determine whether it is possible to experience it anew. After turning one’s attention to the area of phenomenal space in which taste sensations are encountered, one might experience the gradual rebirth of the aftertaste.”<sup>244</sup>

Lycan says that this type of argument rests on a misunderstanding of the HOP theory.

---

<sup>243</sup> Hill 1991, p. 119.

<sup>244</sup> *Ibid.*, p. 121.

“[W]hy cannot the inner-sense theorist grant both that scanning a first-order state can cause a change in the character of that state and that aiming one’s internal monitor at a particular sector of one’s phenomenal field can bring a sensation into existence?”<sup>245</sup>

Lycan does hence not argue that introspection *should* leave its objects unmodified. Inner sense only resembles outer sense in the way Hill thinks it should, if we regard it as a passive form of monitoring, which we should not.<sup>246</sup> Lycan continues:

“I do not think of internal monitoring in this passive, vicarious way; I not only grant, but insist that monitoring often does affect the phenomenal field being scanned.”<sup>247</sup>

Note that what Hill and Lycan are talking about here is distortion in the same sense that was discussed by the psychologists in the early 1900’s. There were worries then that the introspective process somehow would have a distorting effect on the introspected object. According to both Hill and Lycan ordinary perception does not distort the physical object in this sense, but the perceptual experience may be a distortion of the original scene insofar as the intentional object of a perception may not correspond to the physical object which caused the experience. According to Lycan, introspection (AHOP) may be subject to distortion in both senses. The old psychologists again did not, as a rule, consider it possible that the introspective experience should not accurately picture the mental episode it reflected (although the mental state might have changed character as a result of the introspective effort).

---

<sup>245</sup> Lycan 1996, p. 35.

<sup>246</sup> Both Lycan and Hill take for granted that *outer sense* is passive in this sense, and it may seem plausible that physical objects do not change their properties as a result of us perceiving them. But that is only true as long as we restrict the perceptual analogy to the case of visual perception. The tactile sense does not conform to this view. We change things by feeling them as we e.g., often change the structure of the cat’s fur by stroking it.

<sup>247</sup> Lycan 1996, p. 36.

## 6.6 HIGHER ORDER MIS-PERCEPTIONS

There is one especially problematic consequence of the “fallibility thesis” for introspection. The HOP theory is, as it turns out, unable to give a plausible account of unveridical inner perceptions. Karen Neander notices this fact as well. Let us go back to the quote from Lycan where he discusses non-veridical pains. He writes:

“If (as the present hypothesis has it) there is no first-order pain sensation at all but merely a mendacious representation of one, there is no reason to think that all or any of these usual functional effects would indeed ensue. You would be introspecting something that has some of the qualitative aspects of a pain, but important elements would be missing; you might be in the position of the morphine patients, who manifest ‘reactive dissociation,’ saying that they still feel pain as intensely as ever but no longer mind it.”<sup>248</sup>

Neander provides three situations to consider:

- (a) I have a sensory representation of something green and I represent myself as representing something as green.
- (b) I have a sensory representation of something red and I represent myself as representing something as green.
- (c) I have no sensory representation at all and I represent myself as representing something as green.<sup>249</sup>

It seems to follow from Lycan’s theory that there would not be any qualitative difference between these cases from the subject’s point of view, she claims. In each case the subject will have the kind of experience (almost) she normally has when she looks at something green. Neander comments:

“It’s immensely puzzling, to say the least, to see what could even be meant by the claim that what it’s like to experience (a), (b) and (c) is the same, and yet one is phenomenally green, another is phenomenally red, and the third is neither phenomenally red nor green.”<sup>250</sup>

---

<sup>248</sup> Ibid., p. 20.

<sup>249</sup> Neander 1998.

<sup>250</sup> Ibid., p. 420.

Neander concludes that the most obvious reading of Lycan and probably the one he would prefer, is that both (b) and (c) are conscious, but only (b) is qualitative since it involves a sensory representation. I would say that neither (b) nor (c) is conscious. I will return to this in a moment.

Lycan tries to answer to this critique, but admits that it is “very troublesome”.<sup>251</sup> He begins by emphasising that the illusory introspection of a pain would not feel *exactly* like a real pain. What something feels like is only partly constituted by the quale, he says; the other part is determined by the pain’s functional role. The quale is only part of the “overall feel” of a state. In two cases where the first is a higher order perception of a real pain and the other is a higher order hallucination of pain, the introspecteur would notice the difference, Lycan claims, since in the latter case, he would lack any immediate desire to get away from the pain.

But as Neander points out, the case of colour vision is “more motivationally neutral” and is not explainable in the same manner.<sup>252</sup> Lycan tries (unconvincingly) to save things by pointing to the possibility of a red-green colour-blind person who is unable to detect green, but suddenly introspects something as being a green-perception. In this case the colour-blind person would not have exactly the same kind of experience a non-colour-blind would have looking at something green, Lycan says. Probably not, but, in my view, this example does not show anything conclusive about how a non-colour-blind person who mis-introspects something red as green would experience the situation. If she would or would not experience any difference between the two cases.

Lycan finishes:

“My division of phenomenological labour still implies that (b) and (c) are possible, *even though* they would be accompanied by some dissociative features; and in this case differs from that of pain in that there is not so large or obvious a gap between quale and ‘overall feel’. But, remember: Though an introspecteur produces an introspective representation of greenness, that does not amount to its *creating* a sensory green quale (a green quale must be a visual representation of greenness). So it is not that in cases (b) and (c), introspection has substituted or inserted a green quale in place of a red one or none. Nor may Neander assume that ‘what it’s like’ for me in (b) and (c) is the same as what it would be like for me to

---

<sup>251</sup> Lycan 1998, p. 483.

<sup>252</sup> Neander 1998, p. 420.

introspect a green quale under normal circumstances, for higher-order 'what-it's-like'-ness comprehends the overall feel, not just the quale itself."<sup>253</sup>

As far as I can see, this argument does not explain how (c) can be possible since what happens in case (c) will not be accompanied by any qualitative features at all. It is simply not possible to produce a phenomenally conscious representation of green in the absence of a first order state that represents green. The only thing which could perhaps be detected from a subjective point of view, would be a *phenomenal absence* of a motivational force.

I would claim that it follows from Lycan's account of what a conscious state is, that the difference between being in a veridical AHOP or NAHOP and being in an unveridical AHOP or NAHOP is that the latter two would not be like anything at all, from a phenomenological point of view, to be in. A conscious state is a state, which the subject is conscious of. A first order perceptual state is unconscious if the subject is not conscious of it and conscious if the subject is conscious of it. Higher order states are themselves unconscious as long as we are not conscious of them in turn. Lycan writes:

"A state of a subject, or an event occurring within the subject, is a conscious state or event, as opposed to an unconscious or subconscious state or event, iff the subject is aware of being in the state or hosting the event."<sup>254</sup>

So, if a second order representation is only seemingly about a first order state, if the "internal monitor fires without proper cause", *the subject will not be in any conscious state at all*. Let us assume that S is looking at a red apple, and has a perception with the content [red apple]. At the same time S has a higher order (unveridical) perception to the effect that she is in a perceptual state with the content [pink rose]. In that case, she would not know either that she veridically sees a red apple, or that she believes that she sees a pink rose. If she is in a higher order mental state, which is not about a first order mental state, she is not conscious of anything.

---

<sup>253</sup> Lycan 1998, pp. 485-486.

<sup>254</sup> Lycan 1996, p. 3.

Here is an illustration of a first order perception of an apple, according to the HOP theory:

$e \rightarrow \text{apple}$

A veridical higher order perception of an apple:

$\epsilon \rightarrow e \rightarrow \text{apple}$

A “normal” apple-hallucination:

$\epsilon \rightarrow e \rightarrow$

An unconscious apple-hallucination:

$e \rightarrow$

And finally an unconscious higher order hallucination:

$\epsilon \rightarrow$

Since second order states *do not have functional roles*, a second order hallucination will not have any influence on behaviour, verbal or otherwise. The subject will not be able to report about it, neither will she act on the basis of it. A second order representation which is not about a first order state will not make us phenomenally aware of anything, neither will it make us aware of anything in a functional sense.

## 6.7 WATCHING REPRESENTATIONS

One kind of substantial critique against the HOP theory concerns the combination of two important claims which are made by the HOP theorists: that qualia are wide contents of perceptual experiences and that introspection is inner perception. If Lycan holds that qualia are *represented* properties, properties of the objects we perceive, it seems to his critics strange how perceiving the representational vehicles would give rise to an experience of the properties of the represented object.

The HOP theory rests on a materialist foundation. Given that experiences are brain states, it should be those states that we are internally watching. Dretske writes:

“All one can become aware of by scanning (monitoring – choose your favourite word) internal affairs are activities of the nervous system. That, after all, is all that is in there. All that is in the head are the representational vehicles, not the contents, the facts that make these vehicles into thoughts and experiences. You cannot represent a thought or experience *as* a thought or experience, you cannot achieve metarepresentation, by seeing, hearing, smelling or tasting the thought or experience itself. All that experience (in whatever modality) of an experience gives one is a sensory representation of some part of, or process in, the brain. It will not yield a representation of an experience – that part or process in the brain – *as* an experience.”<sup>255</sup>

The naturalist HOP theorists do not try to describe what is going on in higher order perception in neuro-physiological terms (Paul Churchland is an exception). They discuss consciousness in terms of “thoughts”, “perception”, “awareness”, “content” and so on. It is not specified exactly what the metaphysical status of mental representations of the first and the second order is, but this is hardly a fault which the HOP theorists are alone in committing.

Shoemaker claims that perceptual experiences lack perceivable features, since such experiences are completely determined by the *representational relations* they have to features that are external to the mind.

“We can perceive relations between things we perceive; but we wouldn’t perceive these things at all, and so couldn’t perceive relations between them, if they didn’t present themselves as having intrinsic, nonrelational properties.”<sup>256</sup>

Güven Güzeldere makes a similar point: even if we treat the representations involved as *mental* rather than neurophysiological, the problem Dretske discussed still remains. We cannot learn what is being represented by looking at the representation. Güzeldere calls this “the fallacy of the representational divide”. The mistake is to attempt to replace what is being represent-*ed* with that which is the represent-*er*.

---

<sup>255</sup> Dretske 1995, p. 108.

<sup>256</sup> Shoemaker 1996c, p. 205.

“Operating on the basis of this unacknowledged assumption is somewhat like trying out what a stop sign is by studying only the color, shape, material, and mass of the actual sign. Surely one would learn a lot of facts, but expecting to find out in this way what the stop sign *qua* a traffic symbol really is would be a mistake.”<sup>257</sup>

The important issue is this: what features does the representational vehicle have that enables us to become aware of external objects by looking at the representation of these objects? Barry Maund points out that the distinction between properties of the representational vehicle and those of the objects represented is conceptual. It is not ruled out that the vehicle and the object have the same properties or that it is in virtue of having certain properties that a state has the content it does. Maund provides some examples to show that this is sometimes the case.

“Actors on the stage represent characters as doing many things, as having arguments, as fighting, as killing people. Often the actors are pretending or at least simulating. They can represent a killing without actually killing anyone. Nevertheless, there are many occasions where what they represent is done by actually exemplifying what is portrayed. [...] Likewise someone may portray John Malkovich in a movie by being John Malkovich”<sup>258</sup>

Maund anticipates a possible objection: that pictures and actors only represent what they do because people with certain conceptual abilities can interpret them as such. Maund contends that one can insist that perceptual experiences contain both an objective constituent that is sensed and an act of perceptual consciousness of an external object. Furthermore, he says, there is no reason why one could not argue that the sense-datum, or perceptual representation requires conceptual abilities for representation to succeed.

I think Maund is right in holding that it is not impossible that it is in virtue of having certain properties that a state has the content it does, and that it is not ruled out that the “owners” of sensory representations need conceptual abilities in order to interpret these properties. Güzeldere also admits that if the properties of the representational states showed some iconic resemblance relation to that which they represent, there could perhaps be some point in watching them. But he continues, we have no reasons

---

<sup>257</sup> Güzeldere 1997, p. 797.

<sup>258</sup> Maund 2003, pp. 169-170.

whatsoever for thinking that this is the case. According to the HOP theorists' naturalistic assumptions sensory representations are patterns of activation of the central nervous system. It is beyond dispute that these representational vehicles do not "look like" that which they represent.

Gregg Ten Elshof claims that representations *are* suited to be objects of inner sense. He argues that two experiences that are different in virtue of what they represent must be intrinsically different somehow and that this difference could be detected by an inner sense.

"Particular experiences would fail to be well-suited to be the objects of inner-sense if they were not genuine, determinate objects or entities with determinate and discoverable qualities. But a particular experience (say, one which represents there being an orange, round object), like anything else which actually exists, will have determinate features which make it what it is as opposed to something else."<sup>259</sup>

That may be true, but since the HOP theory holds that we become aware of what the world is like by perceiving our perceptual experiences, the discoverable qualities of an experience that represents there being an orange, round object should be orangehood and roundness since that is what we are aware of. But it seems very improbable that an experience is (intrinsically) orange and round, i.e., that the representational vehicle would have these properties.

Armstrong does not describe mental states in terms of their content but in terms of their causal relations to behaviour. He seems to run into the same kind of problems, though. He points out that an apparent difference between inner sense and outer sense is that the objects of the former seem to lack intrinsic perceptible features:

"We would be directly aware of an extraordinarily abstract, and purely *relational*, state of affairs. We would be aware that something of whose non-relational properties we had no direct awareness at all was operating to produce certain behaviour. [...] this is credibly far from the detailed awareness of the intrinsic properties of objects that it yielded by sense-perception."<sup>260</sup>

---

<sup>259</sup> Ten Elshof 2005, p. 57.

<sup>260</sup> Armstrong 1968, p. 96, italics; added.

However, Armstrong then argues that it is a mistake to think that perception must entail awareness of intrinsic properties. Again, this assumption is only adequate as long as we think of vision, he says.

“Suppose I feel a pressure in the small of my back. What am I aware of? It may be that I am aware of no more than this: something I know not what is pressing upon my back. I might say it was something material, but what is a material object in this context except ‘that which is capable of exerting pressure’? I might not even know whether it was something solid, something liquid (such as a jet of water), or something gaseous (a jet of air). My awareness of the object is simply awareness of ‘something which has the relation to me of pressing on me’. Here is a perceptual parallel to the abstract and relational awareness that is attributed to ‘inner sense’.”<sup>261</sup>

While this could make it probable that introspective awareness could be awareness of *causal* relations, as Armstrong thinks, it does not immediately explain how introspective awareness could be perceptual awareness of *intentional* relations.

In subsection 4.2 above I pointed out that Lycan seems ambivalent about whether qualia are representing properties of perceptual vehicles or represented properties. On some occasions Lycan points out that qualia are represented properties of external objects, i.e., they are phenomenal properties of objects. In other passages he seems to claim that qualia are representing properties of the mental vehicle. My guess is that Lycan does want to claim that qualia are represented properties, since that is, after all, what “externalism regarding phenomenal properties”<sup>262</sup>, which Lycan argues for, entails. But since he also wants to maintain that a subject becomes aware of what qualia her perceptions represent by *perceiving* the representations, he sometimes slides over to describing qualia as somehow belonging to the representational vehicle. But, as Güzeldere also points out, picturing qualia as properties of representations only makes sense if those properties show an iconic resemblance to what they stand for.<sup>263</sup>

---

<sup>261</sup> Ibid., p. 97.

<sup>262</sup> Lycan 2001c.

<sup>263</sup> Helge Malmgren (1975) argues that many philosophers have more or less consciously assumed that the relation between a mental act and what it represents is *internal* in the sense that the intrinsic properties of the act uniquely determine the character of the represented object. If an act is connected to its content by such an internal relation, it suffices to look at its intrinsic properties in order to know the content. Iconic resemblance is one example of such relations, but there are other candidates as well. However, I have chosen not to follow up this

Lycan holds that the phenomenology of perception is not entirely reducible to whatever the perceptual experiences represent. He argues that there are modes of presentation under which phenomenal states are given. Could these, as being properties of the representations as such, be what we are aware of in introspection? Modes of presentations do not represent qualia, according to Lycan. Modes of presentations are to be understood as functional roles, and qualia again, are wide contents of perceptual states.

“There is no such thing as representation without a mode of presentation. If a quale is a representatum, then it is represented under a mode of presentation, and modes of presentation may be narrow even when the representational content itself is wide. Indeed many philosophers of mind, myself included, take modes of presentation to be internal causal or functional roles played by representations in question. [...] Remember, the qualia themselves are properties like phenomenal yellowness and redness and greenness, which according to the Representational theory are representata. The mode or guise *under which* redness and greenness are represented in vision are something else again.”<sup>264</sup>

Awareness of the representational vehicle in terms of awareness of its mode of presentation surely does not explain how it is that we are aware of redness, greenness and roundness and so forth, because those properties are not represented by the presentational mode. It may, however, explain why it does not “feel” the same to taste a tomato and to smell it.

## 6.8 SUMMARY OF CHAPTER SIX

The majority of the arguments considered here circle around the relation between introspective awareness and the objects of introspection. Shoemaker argues that the subject is immune to error when it comes to forming introspective beliefs as well as immune to ignorance about her own experiences. Given that he is right, introspection cannot plausibly be compared to perception. The HOP proponents on the other hand contend that introspective judgments are fallible, and that the subject may fail to

---

lead since the philosophers which I discuss all seem to assume that iconic resemblance is the only possibility.

<sup>264</sup> Lycan 2001c, pp. 25-26.

notice features of her own first order experiences (as well as their entire existence). The HOP model is quite dependent on their being contingent relations between the objects of introspection, the introspective process and the introspective beliefs which the subject comes to form as a result of that process. Therefore it is a serious problem that the theory, as it seems, is unable to give a plausible account of unveridical higher order perceptions.

Another major problem for upholding the analogy with perception is that mental states seem to lack perceptible properties. The HOP model seems to entail that the representational vehicles must have the very same qualitative properties as the objects they represent and it is far-fetched to think that mental representations really have those properties we are aware of when we perceive the world.

## 7. Against HOT

Now we will turn to consider some arguments that have been raised against the HOT thesis, the thesis that the subject becomes phenomenally conscious of an object by thinking that she is in a certain state that, in turn, is about that object.

### 7.1 PHENOMENOLOGY

Lycan puts forward as an argument for HOP and against HOT by claiming that experiences of the first order, from a first person point of view, *seem present* to our minds, not merely *represented* by them. The objects of outer perception seem to be immediately present to me in the same way as the objects of introspection, Lycan explains, while thinking about something does not make the intentional object of the thought appear present in the same way. Robert van Gulick elaborates on what “seem present” might mean:

“[T]hinking of an object seems less transparent than perceiving; the means or medium of representation is never fully lost from the experiential view. It is difficult if not impossible to ‘look through’ one’s thoughts as fully as one typically does in perception. The distinction between mental act and object seems phenomenologically more salient in the case of thought and never completely invisible.”<sup>265</sup>

Compare the cases of thinking about the Empire State building and looking at it. When we are thinking about it, it seems obvious that the “vehicle”, that which we are thinking *with* is not the building itself. Assume instead that we are in NYC looking at the building. In this case it does not seem like we are aware of a representation of the building, Van Gulick says. From a phenomenological point of view, the building is *present* in the perceptual case, and only *represented* in the thinking case. Rosenthal replies that the appearance of immediacy that pertains to our inner awareness does not exclude the

---

<sup>265</sup> Van Gulick 2001 p. 289.

possibility that we are aware of our states by thinking about them and not by perceiving them.<sup>266</sup>

This argument against HOT appears a bit odd to me, in what sense do our perceptual *states* feel present to us rather than represented? When I look at the Empire State building, the building seems “present”, alright. When I reflect upon my present perceptual experience the *building*, not any perceptual state, feels equally present. When I think about the building it does, on the other hand, not feel thus present and neither does the building appear present when I try introspecting this thinking process. The phenomenon of “presence” does not seem to support either of the HO theories, but can instead be used as an argument against both of them. The fact that the building feels present when we introspect a perception of it and represented when we introspect a thought about it, might instead show that introspective awareness does not entail *awareness of* the mental state-qua-object in any sense.

## 7.2 ATTENTION

Lycan also holds out that introspection (AHOP) is “attention-like”. We can control our AHOPs with our will, just as in the case of (visual) attention. We can choose whether we want to attend to something or not. This feature is better accommodated by the HOP theory than by the HOT theory, Lycan claims. Rosenthal replies that we can attend to objects of thoughts just as well as we can attend to objects of perceptions.<sup>267</sup>

Lycan seems to think that the fact that introspection feels like (or resembles in a functional sense) visually attending to something points to the conclusion that introspection is perception. Another way of looking at these matters could be that perception and introspection are two species of attention. Can it not be argued that introspecting is *reflecting* upon our thoughts instead of perceiving thoughts, and that both attentive perception and active reflecting can be described as kinds of attention? Lycan argues that the “active searching” which introspection entails cannot accurately be described as “active *reflection*”. He writes:

---

<sup>266</sup> Rosenthal 2002, p. 718.

<sup>267</sup> Rosenthal 2004, p. 21.

“At will, we can selectively attend to environmental region R and detect whatever sensory qualia there are in R. We do not in the same facile way control what regions of or things in the phenomenal field we have thoughts about; the only obvious way in which we control what to have thoughts about is first to attend to a region R and thereby cause ourselves to have thoughts about whatever sensory qualia there is in R. I can *try* to have thoughts about contents of R only by attending to R and detecting qualia there.”<sup>268</sup>

This argument does not seem to prove that introspection is observational, just that it involves an act of attention. The argument does seem to show, though, that introspecting is not merely thinking about our mental states, but at least a kind of thinking which presupposes attention. I will not go into the discussion about what attention is, because the nature of attention is most likely just as obscure as the nature of introspection. I am merely pointing out that just because introspection is a kind of attention it does not necessarily follow that it is a kind of perception.

### 7.3 THE CONTENT OF THOUGHTS AND PERCEPTIONS

If the content of the higher order thoughts is propositional, does this mean that a creature must have linguistic skills in order to have higher order thoughts and thus be conscious? No, a “minimal concept” of self is quite sufficient, Rosenthal argues. To have such a concept simply means that the creature possessing it is able to refer to itself. Such a concept does not need to specify what *kind of thing* the self is. “[I]t need not imply that the self has some special sort of unity, or is a centre of consciousness, or is transparent to itself, or even that it has mental properties,”<sup>269</sup> Rosenthal says.

Similarly, we do not need to suppose any rich conceptual resources in order to be able to refer to our own mental states, Rosenthal claims. An argument for HOP against HOT comes from the suggestion that it seems, on the face of it, impossible to represent a perceptual experience by means of a proposition. The propositional language is just not apt to capture the rich phenomenology of sensory experiences, the story goes. Verbal reports

---

<sup>268</sup> Lycan 2004, p. 105.

<sup>269</sup> Rosenthal 1997, p. 741.

do seldom or never give full justice to a sensory experience, so what are the grounds for suggesting that our introspective higher order representing of perceptual states is propositional? This argument has been put forward by e.g. Lycan (2004), Carruthers (2000) and Byrne (1997). They claim that propositional representations are too *coarse-grained* to accurately represent all aspects of a perceptual experience.

Rosenthal's solution to this problem is that a higher order state refers to sensory states *demonstratively*.<sup>270</sup> What does that mean? Rosenthal writes:

“When a mental state is conscious, it is not simply that we are conscious of the state; we are conscious of being in that state. This places constraints of what the content of these HOTs must be; their content must be that one is, oneself, in that very state.”<sup>271</sup>

But as Byrne points out, taken literally this will simply not do, one can only demonstrate what one is *already* aware of.<sup>272</sup>

Furthermore, the higher order thought must still somehow capture the properties of the first order mental state. *Being aware of the state* must entail that we are aware of what the state is *about* and what phenomenal properties it represents or features. Hence tokening it by a directly referring name will not do this job. Byrne also points this out:

“Take my conscious thought that it's sunny. Because this is what I am consciously thinking, I can report that I'm thinking that it's sunny, and thus express my higher-order thought that I'm thinking that it's sunny. That is, the higher order thought (fully) *describes* the content of the conscious thought. This kind of observation allowed Rosenthal to motivate the higher-order thought account. Now, let us name the mental state that is my thinking that it's sunny, 'Alice'. There simply is no intuitive motivation for the view that my thought that it's sunny is conscious because I have the higher-order thought that I am in Alice, or because I have the higher-order thought that I am in *that* (referring to Alice).”<sup>273</sup>

In the face of this objection, Rosenthal comes up with another suggestion. Instead of one higher order thought that demonstratively refers to a lower order state, we can imagine many higher-order thoughts which jointly

---

<sup>270</sup> Rosenthal 1993.

<sup>271</sup> Rosenthal 1997, p. 741.

<sup>272</sup> Byrne 1997, pp. 117-118.

<sup>273</sup> *Ibid.*, p. 117.

represent the content of a visual experience and which jointly make the entire first order experience conscious, but where some of the thoughts have a *less specified content* than others.<sup>274</sup>

“[W]e may need fewer [higher-order thoughts] than might first appear. [...] For example, the degree of detail we are conscious of in our visual sensations decreases surprisingly rapidly as sensations get farther from the center of our visual field. It is natural to suppose that the content of one’s HOTs becomes correspondingly less specific, and that a progressively smaller number of HOTs will refer to successively larger portions of the visual field.”<sup>275</sup>

It is not clear to me why higher order thoughts with less specific content would demand lesser cognitive resources, than thoughts with a specified content would. To represent the periphery of the visual field propositionally might even take up more (cognitive) space than representing the events in the centre, in so far as the content of that part of the visual experience is undetermined and could for instance take the form of a very long disjunction. (It is yellow, or red, or green, or grey...)

Let us go back to perception for a while. One motive for claiming that we refer to a perceptual feature (e.g., a colour) demonstratively, by naming it “that colour” is that its nature assumedly cannot be captured in words. From a “perceptual point of view”, though, I can still know which colour I am referring to, I can accurately discriminate it from other colour samples, and I may even be able to re-identify the same colour when I encounter it again, expressing this by: “it is that colour again”. I must, however, perceptually attend to the colour before I can refer to it in this way, it seems unlikely that I could attend to a colour *by* demonstratively referring to it. Hence, a non-perceptual belief, expressed by a demonstrative phrase, cannot in itself represent the colour I am talking about. A higher order thought which “is about” a perceptual state cannot successfully refer to that state *only* by describing it as “that state”.<sup>276</sup>

---

<sup>274</sup> Byrne thinks that this solution flies in the face of phenomenology. He says that it does not account for the unity of conscious experience: “I don’t see the pieces of the jigsaw, I see the jigsaw.” (1997, p 118).

<sup>275</sup> Rosenthal 1997, p. 743.

<sup>276</sup> In a later discussion Rosenthal argues that the qualitative character of perceptual states can be captured by *comparative concepts*.

“[T]hinking distinguishes among properties in at least as fine grained a way as does perceiving. And, though we plainly don’t have distinct concepts for all

HOP theories are better apt to explain non-conceptual reference since the introspected object is already perceived. Both Carruthers and Lycan argue that introspective concepts are *recognitional* concepts of experience.<sup>277</sup> These are concepts instantiating the schema: “x is one of that kind”, they are “type-demonstratives”. The idea is that we can form concepts for our experiences that lack connections to any other kind of concepts we might have, functional, physical or intentional. We are able to recognise a certain type of experience as “that experience” each time it occurs.

Rosenthal says that it is unlikely that concepts for qualitative states are recognitional in this sense since all concepts for mental qualities are tied to concepts for other mental qualities as well as to concepts for perceptible properties.

“Rather, our concepts for mental qualities connect in crucial ways with our concepts for the physical properties that those mental qualities enable us to perceive; indeed, our concepts for mental qualities very likely derive from our concepts for perceptible properties.”<sup>278</sup>

I think Rosenthal is wrong about this. As Tye points out<sup>279</sup>; phenomenal concepts do not inform the subject about the metaphysical status of the phenomenal qualities. Nothing in the phenomenal character tells the subject whether the qualities experienced are qualities of *sensa* or qualities of material objects. Loar makes a similar point:

“[I]t is hardly surprising that a recognitional conception of a physical property should discriminate it without analyzing it in scientific terms. Nor should it be surprising that, if there are recognitional concepts that pick out physical properties *not* via contingent modes of presentation, they do not discriminate their references by analyzing them (even implicitly) in scientific terms. Basic recognitional abilities do not depend on or get

---

our conscious mental qualities, that doesn't show that we lack the conceptual resources needed to capture all those qualitative variations, since we can readily capture them using comparative concepts.” (Rosenthal 2004, p. 22.)

There is empirical evidence Rosenthal says, that we are aware of mental qualities by means of comparative concepts. We are aware of more fine-grained differences when they occur together than when they occur one at a time. It is not clear to me why comparative concepts would be of much help here and I have to abstain from discussing the idea further.

<sup>277</sup> They are also discussed by e.g., Loar (1997), Papineau (2002), Sturgeon (2000) and Tye (2000b).

<sup>278</sup> Rosenthal 2004, p. 23.

<sup>279</sup> Tye 2003, p. 38.

triggered by conscious scientific analysis. If phenomenal concepts reflect basic recognitions of internal physical-functional states, they *should* be conceptually independent of theoretical physical-functional descriptions. That is what you expect quite apart from issues concerning physicalism.”<sup>280</sup>

The idea of there being “recognitional concepts for experience” may be able to accommodate the fact that a subject can refer to a perceptual experience that also has non-conceptual content. The HOT account of conceptual thoughts singling out and referring to perceptual experiences seems difficult to uphold.

#### 7.4 THE RELATION BETWEEN A HOT AND A LOS

The relation between a higher order *perceptual* belief and the lower order state (LOS) it is about can be described both in semantic and in causal terms. The higher order percept is *about* the LOS, and the higher order perception is *caused* by the LOS. Rosenthal claims that the requirement of a causal connection between the LOS and the HOT is too strong. “Since mere accompaniment suffices to explain our being conscious of our conscious states, we can remain noncommittal about the causal history of HOTs.”<sup>281</sup> It is also the case that a HOT can be a false belief. Rosenthal claims that it is quite possible for someone to confabulate being in a mental state.

“Because conscious states are sometimes confabulated, the states one is conscious of oneself as being in do not always exist. So we cannot describe a conscious state as a state that bears some actual relation to a HOT. Rather, a state’s being conscious must consist in its being the intentional object of a HOT, the object that the thought seems to be about.”<sup>282</sup>

There is one major problem with this: it is difficult for Rosenthal to explain how thinking about an experience in this special way (HOT) is different from thinking about it in some other way. What if my psychoanalyst tells

---

<sup>280</sup> Loar 1997, p. 602.

<sup>281</sup> Rosenthal 1997, p. 744.

<sup>282</sup> Rosenthal 2002, p. 721.

me that I have an unconscious desire to kill my pet pig and eat him? Is believing that I have this desire because the analyst says so enough to make that state of mine a conscious state? The problem is that if I do not (really) have a desire to kill my pig, me coming to believe that I do, would also seem to make such a desire conscious. We could of course bite the bullet here and simply claim that every mental state I believe that I am in, is a conscious state of mine. Rosenthal seems to express such a view when he claims that we do not need to establish a causal relation between a HOT and its targets “since the targets are simply whatever states the HOTs are about.”<sup>283</sup> Do we not have to put some further demands on this relation in order to exclude these cases of “externally generated” first person beliefs? Rosenthal tries to do so by appealing to phenomenology. He writes: “A state is conscious only if we are conscious of it in a way that *seems* spontaneous and non-inferential. As long as it seems that way, it does not matter how it is caused.”<sup>284</sup>

It seems to me as if Rosenthal in his attempts to distinguish higher order thinking from mere thinking has ended up “half-way” to postulating an introspective process. He claims that the requirement of a causal mechanism for higher order thinking is too strong and at the same time he describes such a process from a phenomenological point of view. Even if there is no introspective or HOT process per se, a HOT differs from some other kind of belief in that it *feels* like it is produced by a special non-inferential belief-generating process.

Furthermore, Rosenthal’s theory of higher order thoughts meets almost the same problems with “false positives” as the HOP theory does. Remember that a conscious state is a state, which the subject is conscious of. A first order perceptual state is unconscious if the subject is not conscious of being in it and conscious if the subject is conscious of being in it. Higher order thoughts are themselves unconscious as long as the subject is not conscious of them in turn. So, if a second order representation is only seemingly about a first order state, the subject will not be in *any* conscious state. Consider the following possibilities:

- (i) S has a true HOT.
- (ii) S has a false HOT.
- (iii) S has a true non-HOT belief.

---

<sup>283</sup> Rosenthal 1997, p. 722.

<sup>284</sup> Rosenthal 2002, p. 721.

- (iv) S has a false non-HOT belief.

Rosenthal claims that S is able to distinguish between on the one hand, (i) and (ii) and, on the other, (iii) and (iv) in virtue of the first two seeming spontaneous, which the latter two are not. It may seem as S can come to believe in a spontaneous way that she has a LOS which she has not and that she can come to believe in a non-spontaneous way that she has a LOS that she has. But, it really follows from Rosenthal's theory that S will not necessarily be conscious of either the false HOT, or the non-HOTs. The HOT theory entails that non-spontaneous beliefs about being in mental states will not make the subject conscious of any such states, if there is none. This may seem like a benefit of the theory and not a problem since it actually rules out the possibility of confusing a non-HOT with a HOT. But since it also rules out the possibility of conscious false HOTs it threatens the independence between states of different orders.

Georges Rey points out that it seems as if Rosenthal puts forward two versions of the HOT theory.<sup>285</sup> One in which he claims that the state which the HOT is about must exist and another where it need not. Rosenthal answers that his hypothesis is the second one and admits that there is a "superficial awkwardness" to this view. He does however not seem to acknowledge the problems that I have described above. He writes:

"[W]hen one ascribes mental states as conscious states or report conscious states as one's own, all that matters is how the subject's mental life appears to the subject. Insofar as we describe a state as being conscious, that state need not actually exist, but can be merely notional."<sup>286</sup>

But if I am right, the subject's mental life will not appear in any way at all to the subject, if the state is "merely notional".

The HOT theory is different from HOP insofar as it pictures introspective awareness as a third order awareness and not as an attentive second order awareness (AHOP). The subject can thus be introspectively aware of false HOTs as well as of non-HOTs posing for HOTs, but she will not from the introspective point of view be able to tell which beliefs are real HOTs and which are not.

---

<sup>285</sup> Rey 2000.

<sup>286</sup> Rosenthal 2000, p. 232.

In my view, the issues discussed in this chapter present serious problem for the HOT thesis. The HOP, unlike the HOT theory, is able to account for how introspective awareness is different from other kinds of awareness we may have of our own states. It is able to do so because it postulates a (causal) mechanism for this special inner awareness. It is also better suited to account for how a higher order state can refer to a state of a lower level. What makes the HOT theory so problematic is not only that it pictures introspective awareness as thoughts about one's being in a certain mental state, but that it assumes that it is such thoughts that make this state a conscious state.

## 7.5 THE CONS OF HOT(D)

Is the dispositional higher order thought theory proposed by Peter Carruthers a plausible candidate? Carruthers' theory avoids some of the problems that the HOT theory runs into. Perceptual experiences are conscious in virtue of being *disposed to cause* HOTs about themselves, the relation between the HOTs and the LOSs is thus secured by this causal mechanism.

There are, however, other problems with this dispositional account of phenomenal consciousness. One serious objection to this theory which has been raised by several philosophers<sup>287</sup> is that *dispositional* thoughts do not seem to be the right kind of phenomenon to explain phenomenal consciousness, since consciousness is something *categorical*. How can something which has not even happened give a perceptual state the unique dimension of *subjectivity*, it is asked.

Carruthers claims that given some kind of "inferential role semantics" such as Millikan's "consumer semantics", this becomes intelligible since the intentional content of a state depends, in part, on what the down-stream consumer systems which can make use of that state *are disposed* to do with it.

"If consumer semantics is assumed, then it is easy to see how mere dispositions can transform contents in the way that dispositionalist HOT theory supposes. For notice that the consumer-system for a given state does not *actually* have to be making use of that state in order for the latter to carry the appropriate content – it just has to be *disposed* to make use of it

---

<sup>287</sup> E.g., Krause and Burghardt 1999, Lyvers 1999, Robinson 1999, Sidel 1999.

should circumstances (and what is going on elsewhere in the cognitive system) demand.”<sup>288</sup>

Experiences that have dual content are conscious, according to this account. Since we do not need to actually think that we are in such a state for such the state to be conscious, it seems as if having this kind of content is enough for the state to qualify as a conscious state. In my view, this still does not explain why a state being disposed to cause a belief is sufficient for the state being phenomenally conscious or exhibiting “experiential subjectivity” in Carruthers’ terms.<sup>289</sup> While modelling the cognitive system as a *consumer system* may explain how the first order perceptual state can carry dual content, it does not in itself explain why a state having this kind of content makes the subject aware of what her experiences of the world are like to her.

A one-level theory accounts for the difference between conscious and non-conscious states by appealing to the functional role these states play. Conscious experiences play a certain role in the cognitive system, and that distinguishes them from unconscious experiences. The crucial point is, of course: how can this functional role constitute the difference between the state being conscious and not being conscious? Carruthers writes:

“It is mysterious why an analog state’s being made available to another set of cognitive systems [...] should suddenly confer on it the properties of phenomenal properties – properties which it did not, by hypothesis, possess prior to being so available.”<sup>290</sup>

But, as Byrne accurately points out, Carruthers is vulnerable to his own objection. Carruthers’ theory also uses a special kind of content plus functional role in order to explain phenomenal consciousness. First order analogue representations of the environment take on a dual content when they are disposed to cause the appropriate HOTS.

Byrne further claims that Carruthers’ theory also entails that there *could be* experiences with dual content that are phenomenally unconscious. He writes:

“Once Carruthers has accustomed us to non-conscious experiences (or, if you insist, ‘quasi-experiences’) with (only) the content *red*, it is hard to see

---

<sup>288</sup> Carruthers 2000b, p. 3.

<sup>289</sup> Carruthers 2000a, p. 129, n.7.

<sup>290</sup> Ibid., p. 170.

why there couldn't be non-conscious experiences with the dual content *red* and *experience of red*. So an experience with dual content won't necessarily be phenomenally conscious, and therefore an extra naturalistic ingredient is required. Functional role is the obvious candidate, but if the earlier objection to FOR theories is correct, this proposal won't work.<sup>291</sup>

In conclusion, it seems as if Carruthers' theory has the same problems as he assumes that the OL theories have; namely that the difference between being and not being phenomenally conscious is accounted for in terms of different kinds of perceptual content. Having dual content seems to be on par with having systemic content or non-conceptual content, as Dretske and Tye prefer to describe this kind of content, in so far as it places the experience in question in special functional position in the cognitive system.

## 7.6 SUMMARY OF CHAPTER SEVEN

The main problem with the HOT theory seems to be that there is no introspective process involved. A subject becomes aware that she is in a certain lower order state by merely thinking that she is. A causal connection between the HOT and the LOS is not necessary in Rosenthal's view, but in the absence of such a process, being aware of a LOS is not sufficient for being aware of it in this special way. The HOT(d) theory provides some kind of mechanism for higher order awareness, since it postulates that a LOS is disposed to cause a HOT about itself, but there is another problem with HOT(d); namely that the existence of higher order thoughts does not in this case explain phenomenal consciousness in so far as a perceptual experience carrying a certain kind of content does not need to actually give rise to a HOT about itself to be conscious.

---

<sup>291</sup> Byrne 2001a, p. 1062

## 8. Against OL

The chief problems with the OL accounts of introspection all basically derive from the OL postulation that introspective awareness is fact-awareness.

### 8.1 INTROSPECTING HALLUCINATIONS

The one-level theories have trouble accounting for introspection of hallucinations. If I just seem to see a green spot, a spot that does not exist “in reality”, the HO theories can simply claim that I am aware of an intentional act, an act that does not represent a real object. Matters become a bit more complicated for the OL theorists since they explicitly deny the possibility of awareness *of* mental acts. Dretske claims that we become conscious *that* we have certain experiences by being conscious of what they are experiences of, but in the case of hallucinations there seems to be nothing the experience is an experience of.<sup>292</sup> So we have two problems when it comes to hallucinations: (i) what are we aware of when we are hallucinating?, and: (ii) how do we introspect a hallucinatory experience? Dretske’s writes:

“I assume [...] that hallucinations are experiences in which one is aware of properties (shapes, colors, movements, etc.) without being *o*-conscious of objects having these properties.”<sup>293</sup>

To be aware of a hallucinatory pumpkin is thus not to be aware of some pumpkin-shaped object, but only to be aware of pumpkin-properties, properties that real pumpkins normally have.

Dretske presents an argument for this analysis by referring to another strange phenomenon where we seem to be aware of properties which do not

---

<sup>292</sup> My discussion of Dretske’s account of hallucinations is limited to what he says in his paper from 1999. He seems to have held another view of what constitutes a perceptual belief in his book from 1995.

<sup>293</sup> Dretske 1999, pp. 106-107.

belong to any object. If you look at a waterfall for a while and then turn your gaze away and look at something else, you will experience movement in the opposite direction from the movement of the water in the fall, but now this movement is not connected to any specific object. It is as if one sees movement but not anything that moves. I would like to point out that these two cases are quite different, however. In the case with the waterfall illusion, the movement *does not even seem* to be a movement of an object – from an experiential point of view. The pumpkin hallucination, on the other hand, is an experience of an orange pumpkin. In this case the colour is *according to our experience* the colour of a pumpkin. There is no physical object causing the experience, but that is a different story.

The notion of “property-awareness” is somewhat difficult to entangle. To be “object-aware” of an object entails, according to Dretske, that there is a real object that causes the experience. What distinguishes a perception from a hallucination is that only the first entails object-awareness in this sense; it is only here that there is a real physical object to be aware of (and which causes my experience). Object-awareness contributes nothing to the phenomenology of experiencing, Dretske says.<sup>294</sup> When we are aware of a real pumpkin we are thus object-aware in the sense that there is a pumpkin causing our experience, but we have no experience of an object from a phenomenological point of view. The object we are aware of seems to be the *x* that carries the properties we are aware of.

Dretske thinks that being aware of pumpkin-properties in the absence of a pumpkin is to be (perceptually) aware of “un-instantiated universals.”<sup>295</sup> This is also, he argues, what happens in the case of “waterfall illusion”. Dretske’s analysis does however not capture the phenomenology of these experiences. In both cases, there is no physical object that presently causes the experience (although the waterfall in some sense is the cause of the illusion). But in the case of the hallucination there *seems to* be an object present, we are not “property-aware” of un-instantiated universals, the orange-ness and bulky-ness seem from the subject’s point of view, to be properties of an object.

To make things even more complicated, some physical “objects” are a bit unusual and do not accord with Dretske’s analysis of object-awareness. We can have veridical perceptions of rainbows or mirror images. These objects are not objects in a usual sense. It would seem to follow from

---

<sup>294</sup> Dretske 1999, p. 108.

<sup>295</sup> Ibid., p. 107.

Dretske's theory that we can not be object-aware of the rainbow, only property-aware of its colours. Does that make perceiving the rainbow a hallucination?

There is also an uncertainty hidden in Dretske's account of fact-awareness. Does fact-awareness entail that we *believe truly* or does it merely entail that we are *thinking that* something is the case? Assume once again that I am looking at the pumpkin and that I *see that* it is orange. Is it necessarily the case that the colour of the pumpkin *is* orange, or does it suffice that I am merely entertaining the thought: "the pumpkin is orange"? Dretske seems to assume that fact-awareness entails knowledge. Throughout the discussion he uses the wording "aware that". We are aware *of* objects (e.g. a pumpkin) and *of* properties (e.g., orangeness), but aware *that* the pumpkin is orange. There is no reasonable interpretation of "being aware that x is the case", which does not entail knowledge, or at least true belief. If a person says that she is aware that today is the 6<sup>th</sup> of September and it is not the 6<sup>th</sup> of September, she is not aware that today is the 6<sup>th</sup> of September. I would say that language does not allow us to be fact-aware of something and be wrong about that something at the same time, and this gives me a strong reason to think that that is what Dretske argues as well.

If this is the case, fact-awareness cannot prevail in the case of the hallucinations, in so far as the pumpkin I believe I see does not exist. Even though I believe that the pumpkin is orange I am not *aware that* the pumpkin is orange, since there is no pumpkin. And this is, of course, also what Dretske wants to argue. But, in a wider sense of being aware, a sense that does not imply true beliefs, I am aware of an orange pumpkin and that the pumpkin *qua intentional object*, is orange. So there seems to be yet another relevant meaning of "awareness", one which Dretske disregards, one in which we ascribe properties to an intentional object.

How do we introspect a hallucination? If we buy Dretske's theory, introspecting a hallucination is no more complicated than introspecting a veridical perception. We are aware of some properties, properties real pumpkins normally have, and we can be aware that we have an experience of a certain *P*-kind, but if I am correct, mere *P*-awareness is not enough in order to accurately describe a hallucinatory experience. It is not, from a phenomenological point of view, the case that I experience "un-instantiated universals".

## 8.2 INTROSPECTING PAINS

Some experiences do not seem to have an object at all, either intentional or physical. Pains are hard to fit into Dretske's schema of objects, their properties and facts about them. If I feel a pain, what object causes my experience, and what object does the experience represent? Dretske argues that feeling of pains do have objects, just like all other experiences. It is a long living mis-comprehension that pains are something purely mental, he says. A pain can be a chemical state in the blood, he proposes. This means that we have an object with properties, just as is the case with every other veridical experience. We can treat pain as Damasio (1994) suggests, Dretske argues, and regard all emotions, feelings and moods as perceptions of chemical, hormonal or visceral states in the body. Then we can analyse the introspections of the experiences of these states in the same manner as we did with the extero-perceptual experiences.

I doubt that this analysis will do. Let us for a moment go back to the different examples of perceptual awareness that Dretske presents in his paper. It is possible, Dretske claims, to be aware of a moving minute hand while not at the same time be aware that it moves, and we can be aware of the minute hand and at the same time being aware that it moves (with or without being aware of the movement itself). How can we translate these examples so as to apply to the case of pains?

“What gives these sensations their phenomenal character, the qualities we use, subjectively, to individuate them, are the properties these experiences are experiences of, the properties (of various parts of the body) that these experiences make us *p*-aware of (irritation, inflammation, time of onset, injury, strain, dimension, intensity, chemical imbalance, and so on).”<sup>296</sup>

It sounds a bit odd to say that a pain experience ( $\epsilon$ ) makes us *p*-aware of chemical imbalance. It seems as if being aware of pain is one thing and being aware of chemical imbalance or inflammation is another. I can, by looking at a sore on my hand become (visually) aware of inflammation. But that does not entail that I am aware of pain. It would follow from Dretske's theory that looking at the sore and feeling the sore would be the same kind of experience since they both constitute awareness of inflammation.

---

<sup>296</sup> Ibid., p. 117.

Furthermore, there are the difficult kinds of pain, pain-feelings where we are unable to find any kind of bodily injury, strain, inflammation, and so on. One example is a felt pain in a phantom limb. What physical object is this experience of pain about? It is hardly the arm or leg where the pain is felt since these do not exist. Due to lack of any other physical object one might feel inclined to refer to some neural state in the brain, but that does not solve our problems here. One could of course say that it is only in the case of phantom limb pain, or referred pains that the proper object of the pain experience is located in the head. But why not let all pain-objects be brain states? And consequently, one could claim the same for sight. My experience of an orange pumpkin can be about an irritation in my retina or some state in the visual centres of the brain.

There might be yet another possibility. We can take the properties we experience pain as having, the phenomenal properties such as perhaps intensity and intolerableness as our starting-points and claim that those are the relevant properties ( $P$ ) of the pain ( $\phi$ ), while the pain-object ( $\theta$ ) is an inflammation or chemical imbalance. Then referred pains and their likes would be on a par with visual hallucinations. They would constitute awareness of pain-properties, but not of pain-objects. This analysis is also suggested by Tye.<sup>297</sup> But since Dretske's analysis of hallucinations is not entirely satisfactory, treating referred pains as hallucinations does not provide a real solution to this problem.

Dretske's analysis of introspection of moods seems to stir up some further problems. We need to fixate which bodily states are relevant for the experience in question. It seems difficult to decide what the relevant properties of a certain mood are. If we take the experienced properties of a particular mood, there are a lot to choose from. Is it the pain in the chest it is about? Or the crawling under the skin or the sting in the eyes? One might say that anxiety is awareness of the conjunction of all present sensations plus some beliefs. But then we would have to say the same for all experiences. Looking at the pumpkin would involve experiences of how the pumpkin smells, how my stomach feels right then, how the wind blowing in my hair feels, and so on. This may seem ridiculous. Naturally the visual experience of a pumpkin should be analysed only in terms of *visual* properties. But how do we know which ones they are? Whether an experienced property is visual or tactile is not given in introspection, according to Dretske.

---

<sup>297</sup> Tye, forthcoming.

One final problem for Dretske: anyone who lacks the concept of experience or the concept under which that which one perceives falls, also lacks the ability to introspect. A young child who is looking at a pentagon and does not know what a pentagon is, might be p-aware of the pentagon's shape. How else explain that the child later can learn to say "pentagon" when she sees a pentagon, Dretske correctly points out. The child will however be unable to introspect her experience of the pentagon.

"Is the child also aware – in *any* sense – of what its experience of these shapes are like, of what it is like to see a pentagon? No. Lacking the concept of a pentagon (not to mention the concept of awareness) the *only* awareness a child has when it sees a pentagon is an awareness of the pentagon and its shape."<sup>298</sup>

Shoemaker points out that it seems natural to speak of pains, itches, tingles and their likes as *objects* of which we are aware.<sup>299</sup> But he, just like Dretske, claims that this is a miscomprehension. Shoemaker can make use of his notion of *appearance properties* to make the story about introspection of pain experiences less troublesome than Dretske's account turned out to be. Pain is a property of a damaged area of the body. It is a phenomenal property and *awareness of pain* can be called *object-awareness*, if we wish, Shoemaker says. Now, introspective awareness of this perceptual state is best thought of as awareness that one is in a somatic state with certain properties. And just as visual experiences have phenomenal character, the somatic experience has phenomenal character, and the latter is fixed by what phenomenal properties it represents. The appearance property of being painful is partly mental and embedded in the perceptual judgment, and the introspective judgment is entailed by the perceptual judgment. The relation between "I feel pain" and "I am aware that I feel pain" is one of conceptual entailment. We will come back to Shoemaker's of "conceptual entailment" in this context, below.

### 8.3 JUSTIFICATION OF INTROSPECTIVE BELIEFS

The HOP theory provides an answer to how our introspective beliefs are justified. Shoemaker and Tye claim that introspective beliefs need no

---

<sup>298</sup> Dretske 1999, p. 119.

<sup>299</sup> Shoemaker 2002.

justification, since they are entailed by *conceptual necessity* by the beliefs they are about. But what is Dretske's view?

### 8.3.1 Dretske's view(s) on justification of introspective beliefs

When it comes to perception, Dretske is a reliabilist. A perceptual belief is justified if it is the result of a reliable causal process. The justification depends only on the reliability of the perceptual process that produces it, that is on the fact that this process leads to a high proportion of true beliefs. Introspection is not, according to Dretske, like perception in this respect. In 1995 Dretske claims that introspective beliefs are formed in a way that resembles what he calls "displaced perception". Introspection is however *more direct* than other forms of displaced perception. This special directness is something which pertains to the information any representational system has about how it, itself, represents the world. It is also *more direct* than the information an external observer has of those representations.

Dretske uses the case of a dog indicating the postman's arrival as an example of displaced perception. What information does the dog-owner (say his name is Sune) need in order to know that the postman has arrived by hearing the dog (say her name is Selma) bark? The barking of Selma gives Sune displaced knowledge of the arrival of the postman iff:

- (i) Sune hears that Selma is barking.
- (ii) Selma's barking indicates the arrival of the postman.
- (iii) Sune believes (ii).

If (i) – (iii) are fulfilled, Sune is justified in believing that the postman has arrived.

In order for Selma's barking (x) to indicate the arrival of the postman (y), x must be dependent on y in such a way that x could not have been the case without the presence of y. This relation has to be "objective", i.e., not depending on some cognitive attitudes on Sune's behalf. Dretske presents two possible ways in which something can be given such an indicator function by natural selection or by learning.

Now, introspection differs from ordinary displaced perception in two important aspects, according to Dretske. He describes the first aspect thus:

“[T]here are two important differences between introspective knowledge and other forms of displaced perception [...] The first point is that although I cannot hear that the mailman has arrived by hearing the dog bark unless the dog is, in fact barking (and I hear it) – unless, that is, I truly represent the the [sic] intermediate fact that ‘tells’ me the postman is here – I can know how I am representing some object without truly representing the intermediate ‘fact’ that provides me with the information about myself. My representation<sub>s</sub> [systemic representation] of  $k$  (as blue, say) does not have to be veridical for this representation to carry information about the target fact, about me, about how I am representing  $k$ .”<sup>300</sup>

My sensory representation of a colour need not be accurate in order for me to get the information I need in order to introspect the experience. I know that I have a certain experience by what it seemingly represents, regardless of whether the representation is veridical or not. But if my representation of the dog barking is not veridical, my representation of the dog will not carry any information about the postman. There will not be any displaced perceptual knowledge in this case.

The second aspect in which introspective knowledge is different from ordinary displaced perception is that the connecting belief in the case of introspective knowledge is infallible, according to Dretske. The connection between Selma’s bark and the postman’s arrival is contingent. Selma may start barking at other people and the connecting belief that Selma’s bark indicates the postman will be false. In the case of introspective beliefs the connecting belief cannot turn out to be false.

“As long as the inference is from what you ‘see’  $k$  to be (whether this is veridical or not) the conclusion must be true: blue must be the way you are representing  $k$ .”<sup>301</sup>

In other words: the connection between what I seem to see to what I take myself to see is not contingent.

These two features make introspective knowledge “less inferential” than other kinds of displaced perception, Dretske claims.

“[I]f this is inferential knowledge, it is a very unusual form of inference. The premises need not be true and the inference cannot fail. As long as

---

<sup>300</sup> Dretske 1995, pp. 60-61.

<sup>301</sup> Ibid., p. 61.

one's conclusion (about how one is representing things) is based on premises concerning how things are 'perceived' to be, one cannot go wrong."<sup>302</sup>

What are the connecting beliefs which are involved in the introspective inferences? Dretske does not tell us. Lycan suggests:

"For me to introspect that the experience is one of blue is for me to represent that it is, but only very derivatively, on the sole basis of the experience's representing the blue object plus the justified belief that the object would not be blue unless the experience (probably) were one of blue."<sup>303</sup>

This is probably a misunderstanding on Lycan's behalf. Dretske argues that the premises concern how things appear to the subject, not how they really are. We will return to this issue below.

Murat Aydede<sup>304</sup> also tries to come up with some plausible candidates for connecting beliefs in this context. He starts from the following suggestions (roughly) of what a perceptual and an introspective belief are:

- 1) a perceptual belief that the object (ball) is blue.
- 2) an introspective belief that I have an experience of a blue ball.

Aydede suggests two candidates for connecting beliefs:

S1: If the ball is blue I am now seeing it as blue.

S2: If I have now come to believe that the ball is blue, I am now having an experience of a blue ball.

S1 is epistemically irresponsible according to Aydede, because it is often false, namely every time I do not look at the ball. What about S2? One problem with S2 is that we have other ways of coming to believe that the ball is blue, than by looking at it. S2 is thus not sufficient to yield the desired introspective knowledge. Aydede makes a third suggestion:

---

<sup>302</sup> Ibid., pp. 61-62.

<sup>303</sup> Lycan 2003, p. 16.

<sup>304</sup> Aydede 2003.

S2\* If I have non-inferentially come to believe that the ball is blue, I am now seeing it as blue.

Aydede points out that this kind of belief might work for the so called proper objects of perception, like “red”, but what about “spherical”? Suppose I have non-inferentially come to believe that the ball is spherical, how can I tell whether I have seen it or felt it? From believing that it is round, I can infer that I have an experience of roundness, but I do not know in which sense modality. Since Dretske stresses that the way an experience “is like” can exhaustively be described by its intentional content, introspection cannot yield knowledge about whether a given perceptual state is visual or tactile if those states have exactly the same content. So S2\* is not an option here, but the following belief is:

S2\*\* If I have non-inferentially come to believe that the ball is round, I am now perceiving it as round.

And we just have to accept the fact that we cannot introspect whether a perception of roundness is visual or tactile.

Both Lycan and Aydede assume that a perceptual belief is a belief about the character of some physical object, but, as I pointed out above, Dretske does not argue here that we infer that an experience is e.g., one of blue from the belief that the *object* we see is blue, but from the belief that the object *seems* blue. There seems to be something strange going on here: Dretske’s perceptual belief is not really a perceptual belief, but an introspective belief. We were supposed, if the analogy with displaced perception should hold, to infer something about our experiences from what they are experiences of. In order to do that we need an introspective belief, a perceptual belief and a connecting belief. But what we have is two introspective beliefs (and one connecting belief). Of course the inference is infallible.

Given the accuracy of Aydede’s analysis, it does not seem as if we could find any working connecting beliefs even if we assume that we proceed from a perceptual belief. To complicate things even further, what about all those perceptual experiences that do not entail any beliefs at all? Say that S is aware of a blue ball, but does not recognise that it is a blue ball, will she then be

able to become aware that she has an experience that represents a blue ball? We will come back to this below.

In a later discussion, Dretske leaves his analogy with displaced perception, instead he uses a new analogy to show how introspective beliefs are justified.

“One does not have to see (be aware of) pictures of *o* (a photograph of a pumpkin, say) in order to be made aware of what these pictures are like. It is sometimes enough to look at what these pictures are pictures *of* (viz., the pumpkins) in order to tell what the picture is like [...] To know what the experience is like, what properties it has, it is enough for the experience[r] to ‘look at’ what the experience is an experience *of*.”<sup>305</sup>

I don’t see the analogy here. The reason we can look at the pumpkin and infer what a photograph of the pumpkin might look like, is that we know how a camera works, how this particular camera is adjusted and where it is placed in relation to the pumpkin. Dretske foresees this objection and tries to meet it:

“All this is true and it is an important disanalogy. It is, in fact the reason why one cannot tell what it is like to be a bat by looking at what the bat ‘looks’ at (perceives) - a moth, say. We don’t know enough about the bat’s ‘camera’ and ‘film’ to tell much about what that bat’s internal ‘pictures’ of the moth are like. But this point, though important, does not upset the usefulness of the analogy. For in the case of experience, the person having the experience, the experienter, looks at objects *with* the ‘camera’ and ‘film’ whose characteristics are (or may be) unknown to third parties. The person sees objects, as it were, *through* the lens and *with* the pre-loaded film - hence, as these objects are represented in his experience of them.”<sup>306</sup>

But Dretske has not provided any argument to the effect that we know more about our own “camera” and “film” than about the bat’s. We know more about e.g., our own feet than about other people’s feet, since we can so to speak, feel them from the inside. But according to Dretske we have no direct knowledge of either the bat’s film or our own, so we know just as much about the bat’s “camera” and “film” as we know about our own, namely that it takes “pictures” of the physical world. The fact that the sensory organs of the bat are very different from ours is not sufficient ground for Dretske’s

---

<sup>305</sup> Dretske 1999, p. 112.

<sup>306</sup> Ibid., p. 123, n. 12.

argument. We do not in either case, have any knowledge about the relation between the “picture” and the world, nor any direct awareness of the pictures as such. Lycan also notes this fact:

”So the suggestion was that for me to introspect, to metarepresent, that  $\epsilon$  has  $\mathfrak{P}$  is for me (a) to represent that  $\epsilon$  has  $\mathfrak{P}$ , (b) on the basis of my seeing that the pumpkin has  $P$ , (c) the former representation being mediated by the justified belief that the pumpkin would not have  $P$  unless  $\epsilon$  had (probably) had  $\mathfrak{P}$ . That can’t be right. I doubt that many people ever have such beliefs as specified in clause (c), much less justified ones – that, e.g., a pumpkin’s being round counterfactually depends on one’s own experience’s being a certain way.”<sup>307</sup>

Here Lycan’s interpretation seems accurate. In this case Dretske is no longer only talking about how things seem to us and what we can infer from that. He has gone over to talk about how things are in themselves, and I agree with Lycan that it seems improbable that people have beliefs like – “the ball would not be blue unless I perceived it as blue.”

### 8.3.2 Tye, Shoemaker and the recent Dretske

Tye thinks that the analogy between on the one hand, displaced perception and, on the other, introspection is useful. He also points out what he takes to be some important dissimilarities between these two cases. He claims that introspective beliefs lack the kind of justification that perceptual beliefs of this kind enjoy. As an example he presents the following case: Tye sets a bomb to go off at 5 p.m. At 5 p.m. he is not in the same location as the bomb. He looks at his watch which says 5:00 and comes to believe that the bomb explodes. The belief that the bomb is exploding at that precise moment is based on a background belief that when the watch reads 5:00 the bomb explodes. Without that belief he would not come to believe that the bomb was exploding. The background belief and the content of the present perception jointly provide a justification for that belief.

In the case of introspection however, there is no corresponding background belief, Tye says. But, introspective beliefs unlike perceptual beliefs based on displaced perception do not need to be justified, he continues. Tye claims that the subject first is *aware of* some external object and then via a *reliable process* becomes *aware that* she has a certain experience.

---

<sup>307</sup> Lycan 1999a, p. 132, n 5.

“If I am aware of certain external qualities, I do not need a background belief to be aware that I am undergoing an experience with a certain phenomenal character, once I introspect. The process is automatic. Introspection of phenomenal character is a *reliable* process that takes awareness *of* external qualities (in the case of perceptual sensations) as input and yields awareness that a state is present with a certain phenomenal character as output.”<sup>308</sup>

Now, Tye further claims that introspective beliefs about experiences with non-conceptual content are generated in the same way as introspective beliefs about perceptual *beliefs* are. The introspective process will, both in the case of experiences with non-conceptual content and in the case of beliefs with conceptual content, make use of so-called “recognitional concepts”.

“One can recognize that one is thinking that water is a liquid, when the only conscious thought one is having is that water is a liquid. In much the same way, we do not have introspective knowledge of phenomenal character by inferring that character from something else. We acquire introspective knowledge of what it is like to have such-and-such an experience or feeling via a reliable process that triggers the application of a suitable phenomenal concept or concepts.”<sup>309</sup>

Maund thinks that this account of going from “awareness-of” an object to “awareness-that” one is in a particular state is problematic.

“The output of the introspection is said to be a state of awareness-that, a state that is conceptual. But what exactly is the content of the state? Do I become aware that I am having an experience with a certain non-conceptual content? That seems bizarre. More plausibly, what I become aware of is the (non-conceptual) content although perhaps I am not aware of it as content. But how exactly is ‘awareness of a certain non-conceptual content’ related to the state of awareness-that? Plausibly, I am aware that there is a tiger before me, which has more features than I can describe – but now it is unclear how this state of awareness-that is different from the state when I do not introspect, but simply see and report what I see.”<sup>310</sup>

Maund is on the whole critical of the usefulness of the concept of “non-conceptual content”. He questions whether a state that entails non-

---

<sup>308</sup> Tye 2002, pp. 145-146.

<sup>309</sup> Tye 2000b p. 53.

<sup>310</sup> Maund 2003, p. 176.

conceptual “awareness of” some properties of objects should be called “awareness” or “experience” at all.

“Is that all it means to say that the state constitutes ‘awareness of the qualities’: that it carries this content? That surely is not awareness in anything like a normal sense. Nor is it remotely like an experience. Why call it awareness or experience then?”<sup>311</sup>

This is a complicated issue in its own right. The one level accounts of both Dretske and Tye rest on the assumption that mental states that “carry” non-conceptual content are experiences, and that having this kind of content is sufficient for being an experience. This view has been contested by several philosophers, e.g., by John Heil.<sup>312</sup> Maund’s point seems to be that since an “experience” of this kind does not even qualify as an experience, what Tye then describes as introspective awareness is instead perceptual awareness.

I agree with Maund that there is something strange about Tye’s story. I do however not think that experiences without beliefs necessarily are non-options given a representational theory of perception. It might still be possible to give an account of experiences representing properties of external objects, while not representing them *as* properties of a certain kind. But if a subject looks at a papaya and does not recognise it as a papaya, it does not seem possible for her to form an introspective belief with the content “I see a *papaya*.” The introspective beliefs that experiences of this kind can possibly produce are of the kind: “I have an experience with a certain phenomenal character”. This however requires that the subject has the concept of experience.

Tye’s theory of how we arrive at introspective beliefs is very similar to that of Gareth Evans’. Evans describes the introspective process thus:

“Although the subject’s judgements are *based upon* his experience (i.e. upon the unconceptualized information available to him), his judgements are not *about* the informational state. The process of conceptualization or judgement takes the subject from his being in one kind of informational state (with a content of a certain kind, namely, non-conceptual content) to his being in another kind of cognitive state (with a content of a different kind, namely conceptual content.)”<sup>313</sup>

---

<sup>311</sup> Maund 2002/2003.

<sup>312</sup> Heil 1991.

<sup>313</sup> Evans 1982, p. 227.

What the subject does, in order to arrive at an introspective judgment, is basically to go through the same procedure as that which precedes a perceptual judgment. In the case of introspection, however, he excludes any knowledge he has which is of “an extraneous kind” and instead “prefixes” the result with “it seems to me as though...”.<sup>314</sup> I understand this claim like this: assume that a person sees that there is a green tractor in the field. She will then be able to judge both; “there is a green tractor in the field” and; “it seems to me as though there is a green tractor in the field”. In the latter case she disregards any knowledge or information she has about the external world (including the tractor) and “prefixes” the result: “it seems to me as though...”

Evans points out that the subject must understand the content of his judgment “it seems to me as though *p*...” . Specifically, this understanding must enable him to differentiate between this judgment and that of “possibly *p*”. In a sense, he must have the concept of experience just as self-ascribing beliefs requires “possession of the psychological concept expressed by ‘ $\xi$  believes that *p*’.”<sup>315</sup> Evans writes:

The procedure I have described, of re-using the conceptual skills which one uses in order to make judgements about the world, is not by itself enough for the capacity to ascribe experiences to oneself.”<sup>316</sup>

Dretske seems to have changed his view about the possibility of justification of introspective beliefs. In a late discussion<sup>317</sup> which concerns beliefs and not perceptions, he claims that our beliefs about our attitudes’ contents simply and automatically inherit those contents. Dretske does however not think, as Tye does, that perceptual awareness which does not entail concepts can be introspected.<sup>318</sup> And Dretske, just as Evans, also thinks that the subject needs the concept of experience in order to be able to make introspective judgments. Shoemaker sides with Evans and Dretske on the former issue. He claims that it is only when we are in a position to pass a perceptual judgment that an object has a certain property that we can pass

---

<sup>314</sup> Ibid., p. 228.

<sup>315</sup> Ibid., p. 226.

<sup>316</sup> Ibid., p. 230.

<sup>317</sup> Dretske 2003b.

<sup>318</sup> Dretske 1999, p. 119.

the introspective judgment that our experience represents the property. In my view, this is an accurate conclusion given the premises.

Shoemaker further claims that introspective judgments are *entailed* by perceptual ones.

“The relation of ‘That looks<sub>p</sub> blue to me’ to ‘I am having an experience as of something that looks<sub>p</sub> blue’ seems to be one of conceptual entailment. So the introspective awareness expressed by the second proposition seems implicit in the perceptual awareness expressed by the first.”<sup>319</sup>

I do not quite understand the idea of introspective beliefs being conceptually entailed by perceptual beliefs. While the introspective judgment may be *implicit* in the perceptual awareness, the introspective judgment is not, in my view, *conceptually* entailed by the perceptual judgment, unless one assumes that the concept of experience is entailed by statements about how things appear. Then again, if it is assumed that the concept of experience is entailed by the content of perceptual judgments, the account becomes rather trivial. On the other hand, this seems to be what Evans wants to point out, his basic claim being that introspective knowledge is not very different from perceptual knowledge. In order to arrive at an introspective belief, we do not start examining the experience, we take one more look at the world, he says.

#### 8.4 SUMMARY OF CHAPTER EIGHT

Neither the HOT theory nor Dretske can provide an answer to how our introspective beliefs are justified. The HOP theory may thus be motivated on epistemological grounds. The HOP theory entails that introspective beliefs are justified by our object-awareness of our mental states. Lycan writes:

“Our awareness of our own mental states *justifies* our beliefs about them. Indeed, my only justification for believing that I now have a slight ache in my tailbone and that I am hearing sounds as of a car pulling into the driveway is that I am aware of both states. Active introspection justifies our beliefs about our mental states even more strongly, though by no means infallibly. This is just what we should expect if awareness is a perception-like affair. By contrast, merely having an assertoric thought to

---

<sup>319</sup> Shoemaker 2002, p. 466.

the effect that one is in state M by itself does nothing to justify the belief that one is in M, and having a metathought about that thought adds no justification either.”<sup>320</sup>

Van Gulick points out that the HOT theory can not be fit into a *reliabilist* scenario either.<sup>321</sup> The having of thoughts cannot in itself be considered as a reliable information channel.

The OL theory of conceptual entailment is not crystal clear either. How is it that a belief like “I have such and such an experience” is entailed by the belief “the physical object is such and such”? Dretske’s and Tye’s theories also have difficulties accounting for introspections of perceptual experiences that do not at all entail beliefs. Experiences that only encompass object-awareness cannot be introspected, Dretske accurately concludes. Whenever one is in a position to make a perceptual judgment that a thing has a certain property, one is in the position to make an introspective judgment that one’s experience represents that property (given that one has the required concepts for such judgments). When one is not in such a position, one is not in a position to pass an introspective judgment either.

---

<sup>320</sup> Lycan 2004, p. 107.

<sup>321</sup> Van Gulick 2001.

## 9. Conclusions

Several arguments have been raised against each of the theories discussed here. Below I will briefly summarise the problems for each theory that seem, to me, most crucial and explain why neither the HO theories nor the OL ones can, in my view, give satisfactory accounts of the phenomenon of introspection.

### 9.1 HOT

The HOT theory entails that conscious states are states a conscious being is conscious of being in. Such a being becomes conscious of a certain mental state of hers by thinking about it, or more specifically, by thinking that she is in this particular state. The two primary problems for this kind of theory are: (i) accounting for how a subject can refer to a mental state of hers that she is not previously conscious of, and (ii) providing an answer to what it is that constitutes the difference between having a higher order thought about something and thinking about it in some other way.

Rosenthal argues that the subject becomes aware of a certain state by having a thought that refers to the state demonstratively. Byrne's point here, which I think is accurate, is that one cannot refer to something one is not already aware of in some sense. It is not the case that we can refer to things we are unaware of (except perhaps accidentally). The HOP theory has an advantage here since it provides an answer to the question how we can become aware of a mental state. The answer is: by perceiving it.

In the case of ordinary perception, we seem to sometimes become aware of something as a result of a voluntary act on our behalf. For example, suppose a friend tells me: "Look at that deer!" I say: "What deer?" And he replies: "It is out there in the field, close to the trees." I try to see the deer by focusing on the part of my visual field which I think is the relevant region, and *if* I discover the deer I can refer to it: "I see *it*!" Here I have voluntarily tried to become aware of something, but this would not have been possible if there were no visual field or something corresponding to such a field, to

attend to. Suppose my (philosophical) friend would have said: “Try to think demonstratively about the deer I see right now out there in the field!” Would I be able to do so? Remember that I have to think about *that* deer, not just any deer, *as* “that deer”, if the analogy with Rosenthal’s theory shall hold.

The second serious problem, which is closely related to the first, is that the HOT theory cannot explain the difference between being “HOT-aware” of being in a mental state and being aware of the mental state in some other, less HOT, way. Rosenthal tries to account for this crucial difference by appealing to the phenomenology of having HOTs. He claims that having a HOT with the content “I am in this state” will make the awareness of the state seem “non-inferential” from the subject’s point of view. It is hard to see the point of this suggestion. Why maintaining that there is a special phenomenology connected to these cases that does not reflect anything on either a functional or an intentional level, when the point of the whole theory is to explain phenomenal consciousness by appealing to intentionality?

## 9.2 HOT(D)

Carruthers’ theory avoids the problems just mentioned. Since conscious experiences are disposed to *cause* HOTs about themselves, the relation between the HOT and the LOS is secured by this causal mechanism.

It has been argued that a major problem with this theory is the claim that being disposed to cause a HOT is sufficient for a state being conscious. The only way I am able to understand this claim, is if one reads Carruthers’ theory as a version of an OL theory. Conscious experiences are according to Carruthers experiences with a certain type of content and having this type of content entails that the cognitive system is *disposed* to make use of it should circumstances demand. Having dual content seems functionally equivalent with having systemic content or non-conceptual content, as Dretske and Tye prefer to describe this content, in so far as it places the experience in question in a special functional position in the cognitive system.

One could, of course, overlook this startling hypothesis and focus on the idea of dual content which seems rather innocent in itself. Conscious experiences are experiences that have the content [the object is such and such] and the content [the experience is such and such]. This seems to explain introspective awareness quite easily. One simply shifts focus from the “worldly content” to the “experiential content” of the experience in

question. This would, in fact, make Carruthers' theory very similar to Brentano's. Since Carruthers also claims that qualia are narrow contents of perceptual experiences, it is less difficult for him to explain how the subject can become aware of what her experiences are like to her, by being aware of the experience as such (qua representational vehicle).

### 9.3 HOP

The higher order perceptual theory seems to have some advantages over the higher order thought theory since it can give a more plausible account of how we become aware of our own mental states. According to the perceptual metaphor, we become higher order aware of being in a state by "perceptually attending" to that state. This theory postulates a causal mechanism for higher order awareness, thus distinguishing it from other ways we can be aware of our mental states. If, for example, a subject thinks she has an unconscious desire, this is not sufficient for making the desire conscious. It is only when she becomes aware of the desire in this special HOP way that the desire becomes conscious. The HOP theory can also provide a better explanation of how a higher order representation can be *about* a representation of a lower order than the HOT theory is able to do. We do not, according to the HOP theory, become conscious of a state by referring to it, but are able to refer to it because we are conscious of it.

There are however other problematic consequences of the HOP theory. One of these problems stems from the attempt to combine phenomenal externalism with the perceptual model of inner awareness. Qualia are according to Lycan properties of external objects. This means, of course, that qualia are not intrinsic properties of perceptual states. Now, Lycan also claims that we become *aware of* qualia by attending, in a way that resembles perceptual attention, to the states that represent qualia. Several philosophers have pointed out the implausibility of this thesis. Perceptual awareness seems, first and foremost, to be awareness of intrinsic properties of objects. If perceptual states lack (relevant) intrinsic properties, it seems unjustified to claim that the subject's awareness of her mental states is similar to her awareness of external objects. Shoemaker, e.g., claims that given the fact that perceptual states are determined by their content and not by their intrinsic properties, it follows that we do not perceive these states, in any plausible interpretation of the term "perceive". Shoemaker, as well as Dretske and Tye

claim that the subject's awareness of her own mental states should instead be conceived of as a kind of "fact-awareness" unmediated by "object-awareness" of those states. The subject becomes *aware that* she is in a particular state, but she is not *aware of* that state. The phenomenal character of introspecting cannot be explained by a direct perception-like awareness of mental states, since these states do not themselves have any of the properties we are aware of while being in them.

In my view, Lycan on occasion slides over from his explicit thesis that qualia are represented properties of external objects to describing them as representing properties of experiences. As in the following quote from Lycan:

"[I]ntrospectings do not have qualia even though they represent first-order states *as having them*."<sup>322</sup>

It is difficult to understand what Lycan really says in the quote above. He explicitly denies that introspectings *have* qualia. However, introspectings represent first order states *as having* qualia. So while he in other places denies that first order states have qualia, he here claims that they *seem to have qualia* from a subjective point of view. If qualia were representing properties of experiences, the theory would be easier to understand, in my view, even if it would not necessarily make it more plausible. Then the story would go something like the following: A perceptual experience represents some property of an external object as being e.g., blue (disregarding for a moment how hard it is to account for blue as a physical property). The experience of blue differs from an experience of red in that the first experience has a certain quale that represents blue, while the other experience has another quale that represents red. Since a mental state is only conscious if the subject is conscious of that state, the quale that represents blue may not be known by the subject and she will in that case not be aware of what it is like to experience blue. If she attends to her visual state and thus becomes aware of it she will become aware of the blue-representing-quale and of what it is like to have an experience of blue. This is however not what happens according to Lycan. The quale blue is a property represented by the perceptual state, it is part of its wide content. So, when the subject attends ("perceptually") to her visual state she will instead become aware of a phenomenal property of an external object. To me it seems as if Lycan tries to hide the implausibility of this thesis, by claiming that, to the subject, it seems as if her first order

---

<sup>322</sup> Lycan 2003, p. 26, italics; added.

states have qualia, but he has not provided an explanation to why it would seem that way, when it is not like that in reality.

The pre-externalist, pre-representationalist version of the inner perception theory (advocated by e.g., James and Brentano) does not encounter the problems HOP struggles with since this theory presupposes that experiences have intrinsic properties in a straightforward sense. What any given perception feels like is according to these thinkers (in my modern reconstruction) determined by its intrinsic qualities. By “looking at” or attending to our experiences we can therefore become aware of what they are like. The HOP theory may be regarded as an attempt to keep the perceptual metaphor for the introspective process, while changing the putative objects of introspection. As we have seen, this does not work.

Is there any other possibility? Maybe perceptual awareness does not require awareness of intrinsic properties of objects? Dretske suggests that Lycan could be interpreted as holding that the introspective scanner yields *knowledge that* we have experiences without making us *aware of* the experiences themselves.<sup>323</sup> Lycan strongly objects to this.

“If the scanner does not ever make me aware of my experiences themselves, by what other means does it furnish me with the knowledge that I have them?”<sup>324</sup>

This is a good question. To be sure, we can gain a serious amount of awareness of facts without being aware of the objects to which these facts obtain. We can, for instance, become aware of how much gasoline there is left in the tank, by looking at the gas meter and not by looking into the gas tank itself. But this idea does not seem to work in relation to the HOP model. If the introspective scanner makes the subject aware that she has a certain experience by making her directly aware of some *other* object and not of the experience itself, what could that other object be? It cannot be the object that the experience is an experience of, as Dretske and Tye claim. That would imply that the “introspective scanner” has to monitor the external world, which in turn would turn the introspective scanner into an ordinary perceptual scanner.

It needs to be added that Lycan’s theory also entails properties of perceptual states that we are higher order perceptually aware of in a more

---

<sup>323</sup> Dretske 2003a, pp. 8- 9.

<sup>324</sup> Lycan 2003, p. 25.

straightforward sense. One experience differs from another in so far as they are presented in different *modes*. But even if the mode of presentation could perhaps be regarded as an intrinsic property of the experience, this does not solve the HOP problem of qualia. Lycan is careful to point out that modes of presentation are not “candidates for qualitative content”.<sup>325</sup>

Another problem with the HOP theory has to do with introspective misrepresentation. Just as we can misperceive, we can innerly misperceive, Lycan holds (and should hold if he wants to maintain the analogy with perception). A conscious experience involves, according to Lycan, two kinds of representations: a first-order representation of an external object and a second-order representation of that representation. A felt pain, e.g., involves a representation of some bodily damage and a second-order representation of the first representation. First order representations are responsible for determining the quality of a sensation (by means of their wide contents), while the second order representations determine whether the sensation is conscious or not. From these facts follows, Neander points out, that it is difficult to describe what a subject’s *conscious experience* would be like in cases where the second order representation is a mis-representation. Suppose, e.g., that the first order (sensory) representation represents something as green, and the second order representation represents the subject as representing something as red. While Lycan argues that the conscious experience of a real feeling of pain would differ from the experience of an imaginary pain feeling in that there is a motivational force only felt in the first case, it is difficult to see how he could do so in the case of colour vision. He must somehow claim that there are some specific feelings combined with perceiving green which do not have to do with how the colour appears to the subject and which are absent in the perceiving of red, but there is no phenomenological evidence that speaks in favour of such a claim. The argument about the colour-blind person who is unable to detect green, but suddenly introspects something as being a green-perception, is not very telling, in my view.

If I am right, the veridical and the unveridical higher order representation of green would differ from one another in so far as the latter would *not feel like anything* at all to the subject. The higher-order state will be unconscious in so far as the subject is not conscious of *that* state, and the lower-order state will be unconscious since the higher-order-state is not about *it*. Neander seems to think that there would be something it is like to have an unveridical higher order perception. She says that in the event of a

---

<sup>325</sup> Lycan 1998, p. 480.

person having a sensory representation of something red while representing herself as representing something as green, this would be phenomenally green to the person. I agree with Neander that this is all “immensely puzzling”. According to Lycan first order states represent qualia, second order states represent first order states. Being in a second order state will make the subject aware of the first order state, and therefore, in turn aware of the qualia it represents. In the alleged case where there is no first order state at all, but where the subject nevertheless has a higher order representation of a perception of green, nothing will be conscious unless of course, the second-order state somehow “creates” a first order state that is like something to be aware of. It is not enough that the second order state has an intentional object, any intentional object, it must be about a real first order state. This reconstruction of the HOP theory is probably not something Lycan would accept, but it is difficult to find another one which makes more sense of the fallibility thesis.

#### 9.4 DRETSKE

Dretske is well aware of several of the problems with the HOP thesis that I have pointed out here and presents an alternative theory for introspection, still based on the premise that qualia are part of the wide content of perceptual experiences. As we have seen, this theory is not without its drawbacks either. Dretske is well aware of some of these, such as that a subject needs certain concepts in order to form introspective beliefs. She needs the concepts of experience and of representation, and she also has to be able to describe the content of her perceptions in conceptual terms in order to be able to introspect these perceptions. He also points out that it follows from his strong representationalist theory and phenomenal externalism that a subject cannot know, introspectively, whether she sees or feels a specific shape, neither can she know that it is she who does the perceiving.

If we follow Dretske’s discussions over time he seems to limit the scope of introspective knowledge for each article he writes. In 2003 he claims that we cannot learn through introspection *that we have a mind*. First person knowledge only extends to knowledge about *what* we think or perceive and not to *that* we think and perceive it. He also seems to accept that first person

beliefs cannot be justified, since he claims that our beliefs about our the contents of our attitudes simply and automatically inherit those contents.<sup>326</sup>

This latter conclusion may be the result of the fact that he was not able to give a plausible account of how introspective beliefs are justified. In 1995 he compares introspective knowledge with so called “displaced perception”. A person can come to believe that something is the case, not by directly perceiving the objects involved, but by perceiving something else. But in order for such beliefs to be justified the subject also needs some connecting beliefs, beliefs about how that which she sees can say something about the circumstances he becomes aware of. Remember the case with Sune and his dog. Sune is justified in believing that the postman has arrived when he hears Selma barking, if (i) Selma’s barking indicates the arrival of the postman and (ii) Sune also believes that this relation holds. Dretske has however not been able to provide any corresponding connecting beliefs for introspective beliefs. If he would maintain, as Lycan thinks he does, that the connecting beliefs in the case of introspection were of the kind: “the external object would not be blue unless the experience were one of blue”, he would hold something very implausible. If he on the other hand claims, as I read him, that the belief is of the form: “the external object would not *seem* blue unless my experience were one of blue”, the analysis seems trivial since we “infer” one introspective belief from another introspective belief.

Dretske does not seem to appreciate the fact that his analysis of introspection does not work very well for introspection of pains. He claims that we become aware that a certain experience of ours have certain properties by being directly aware of the properties the experience is an experience of; in the case of pain we are aware of properties such as chemical imbalance and inflammation. In my view, it seems as if being aware of pain is one thing and being aware of chemical imbalance or inflammation is another. I can by e.g., looking at a sore on my hand become (visually) aware of inflammation. From a phenomenological point of view, these cases are different. By looking at an inflammation and by feeling an inflammation we become aware of quite different properties and, according to Dretske, property-awareness determines the phenomenology of experiencing. If Dretske wants to maintain this thesis it seems as if he will have to find some other (objective) pain-properties of which we are aware when feeling pain.

It is also difficult to account for hallucinations, given Dretske’s schema of objects, properties and facts. Dretske puts forward an attempted solution

---

<sup>326</sup> Dretske 2003b.

to how awareness of hallucinations should be analysed, which is not in my view entirely satisfactory. If a subject has a hallucination of a pumpkin, she will not be object-aware of a pumpkin, since object-awareness is a causal relation between real objects and experiences, Dretske claims. She will however be “property-aware”. In this case it means that she will be aware of properties that do not belong to any object. He compares these cases with the “waterfall illusion”. In the latter case one experiences movement, but nothing that moves. From a phenomenological point of view this is not what happens in the case of hallucinations. If I hallucinate a pumpkin (perhaps without being aware that it is a hallucination), I will, in an important sense, be aware of an orange *object*. Since object-awareness according to Dretske entails standing in a certain kind of causal relation to a physical object, he is unable to account for the phenomenological difference between hallucinating a pumpkin and “illusinating”<sup>327</sup> a movement that does not pertain to any moving object.

If we disregard the last objection to Dretske’s theory (since the problem of explaining awareness of hallucinatory “objects” is not particularly a problem for a theory of introspection), it follows from the other points raised here that introspective awareness, given Dretske’s story, is extremely limited in scope. We can be introspectively aware that we have an experience with a certain content, but only given that the experience involves fact-awareness of the object perceived, that the subject also has the appropriate concepts for forming introspective beliefs (such as the concept of experience) and that the experience is not one of pain. And this is all! We cannot by means of introspection become aware of whether we see something or touch something (in any interpretation of the term “aware”). Nor can we become “self-aware” in the sense of becoming aware that it is we who are perceiving or becoming aware that this is what the world is like “to us”, or whatever formulation one may use to capture this illusive phenomenon. The HOP theory is able to account for all these examples, but the price is, as I have pointed out, too high.

---

<sup>327</sup> - having an illusionary experience of something or experience a illusionary object.

## 9.5 TYE

Since Tye's theory is in many aspects similar to Dretske's, he runs into the same problems trying to account for pains and hallucinations. There is, however, one important difference between the two theorists; while Dretske holds that the subject has to be "property-aware" of external objects in order to be able to introspect her perceptions, Tye thinks he can account for introspective awareness of mere "awareness of" by appealing to the workings of an "automatic process". Dretske argues, accurately, that if the subject does not have any beliefs about what it is she sees, she will not be able to form any relevant introspective beliefs either. As he points out, one cannot be aware that one sees a pentagon unless one masters the concept of pentagon. Consequently, one cannot be aware that one has an experience that represents a pentagon if one does not have that concept. Introspective awareness comprises fact-awareness of one's experiences mediated by property-awareness of external objects, according to Dretske.

Tye on the other hand thinks that, from being in a state with nonconceptual content, a subject can go to being conceptually aware that she is in a certain mental state, but as Maund has pointed out, it is difficult to see how this could be done. Assume that a person is looking at a pentagon without recognising it as a pentagon. Her perceptual experience will, according to the representational thesis, still represent the pentagon, but non-conceptually, according to e.g., Tye. The content of introspective states on the other hand, is always conceptual, according to both Tye and Dretske. So how does the reliable process Tye postulates work?

If Tye is assuming what Evans also holds, accounting for how we come to arrive at introspective judgments is not more difficult than explaining how we can on the basis of sensory experiences form perceptual beliefs about what we perceive. But, according to Evans, it still entails that we master the concept of experience and that we are able to distinguish between "it seems to me as though p..." and "possibly p". That is, between a belief about how we experience the world and a tentative belief about how something in the world is.

## 9.6 SHOEMAKER

Shoemaker is not a strict phenomenal externalist. He tries to strike the golden mean and argues that perceptual experiences represent objective properties of external objects as well as phenomenal properties. In so far as experiences represent phenomenal properties he can fairly easily explain how introspective awareness works. The states of affairs represented by our experiences are partly mental, he claims, which means that part of the content of the perceptual experience is “closely related” to the content of an introspective awareness of this experience that the subject might have, Shoemaker claims. Objective properties are however also represented. Shoemaker claims that objects may appear to have properties objectively, and they may also appear to have properties merely phenomenally.

Now, Tye argues that Shoemaker seems to evoke an “appearance/reality distinction for the colours.” In my view, Shoemaker here points out an important fact about our experiences. It is sometimes the case that an object appears to have a colour we do not think the object really has. We can hence perceive something to have a (phenomenal) colour we do not believe is an (objective) property of the object. Sometimes our beliefs about what it is we see are actually given in the experience as such. As when I see that the surface of the wall is white, but looks red because it is lit up by a red spotlight, or when I see that the coin on the table is round, but looks elliptical. Compare these cases to the case when we know something about some external object while not in any sense “seeing it”. I may for instance know that the two lines in the Müller-Lyer illusion are of equal length, yet I am unable to see it. Whether Shoemaker actually succeeds in accounting for how our perception can represent both objective and phenomenal properties, I am not prepared to answer here, but his basic point is no doubt valid.

He leaves me a bit curious about how exactly he pictures introspective awareness, though. To me it seems as if Shoemaker evokes an appearance/reality distinction for *phenomenal* colours. He writes, e.g.,:

“Let it be that I am perceptually aware that something has the occurrent appearance property of looking<sub>p</sub> blue, and am introspectively aware that I am ‘appeared<sub>p</sub> blue to’, i.e., that I am having an experience as of something that looks<sub>p</sub> blue. The relation of ‘That looks<sub>p</sub> blue to me’ to ‘I

am having an experience as of something that looks<sub>p</sub> blue' seems to be one of conceptual entailment."<sup>328</sup>

What does the phrase "having an experience *as of* something that looks phenomenally blue" entail? It seems as if the subject somehow doubts whether the experience is *really* one of phenomenally blue, or just one that appears like an experience of phenomenally blue. It thus seems as if Shoemaker claims that a subject can question how things appear to her, something he has vigorously criticised elsewhere. Or is "I am having an experience as of something that looks phenomenally blue" just a synonym for "Something looks phenomenally blue to me"? If so, Shoemaker's analysis of introspection threatens to become completely trivial. I do not know which interpretation should be preferred.

## 9.7 FINAL REMARKS

In my view, this discussion has shown two things. Firstly, theories that claim that phenomenal consciousness requires some kind of higher order representations of perceptual states do not, as a rule, work. Secondly, any theory that maintains phenomenal externalism will have great difficulties accounting for introspection. Rosenthal does not think that phenomenal character of experiences can be exhausted by the representational content of these states, but he, on the other hand, holds a most implausible theory of higher order awareness. Lycan's higher order theory would make sense given Rosenthal's view of phenomenal character, and perhaps even if he claimed that qualia are narrow as Carruthers does. Instead Lycan argues that qualia are wide contents of experiences and so his theory ends up being implausible as well.

It should be added that in the case of some of the HO theories, it is not the account of introspection specifically that is problematic, but the theory of phenomenal consciousness in general. The theories of introspection put forward by both Carruthers and Rosenthal are not especially awkward in themselves. Then again, a theory of introspective awareness that does not in any way propose an answer to the more general question of the nature of consciousness is not very informative. The situation is different for Lycan;

---

<sup>328</sup> Shoemaker 2002, pp. 465-466.

even if he rejected the higher order theory of consciousness, his account of introspection as inner perception would still be questionable given that qualia are represented properties of objects.

In my view, the one-level theories make slightly more sense in general, and especially in Dretske's version, since he bites the bullet of phenomenal externalism and asserts that the subject has only introspective access to the representational content of her experiences. That seems to accord with what representationalism and phenomenal externalism entail. But if one thinks that introspecting reveals something more than the representational content of our experiences, Dretske's theory is not very convincing. Is it not the fact that we can tell not just what we believe but also how strongly we believe this? And can we not tell whether we see something or feel it and also tell how it makes us feel to perceive something? Does it e.g., make us feel sad, worried or full of expectations? Dretske's view of introspection does not present any account of such introspectings.

Shoemaker, finally, seems to acknowledge most of the difficulties I have discussed here, which is probably the reason why he advocates his "externalist-internalist" version of representationalism. Could it not even be the case that features such as pitiful and pleasant were phenomenal properties of physical objects? However, it is difficult to see how his "entailment" account can solve the problem of introspective knowledge.

One might conclude that introspective awareness as it once was thought of does not exist or that either representationalism or phenomenal externalism is a false theory. However, I want to be more careful and although I claim to have shown that these particular attempts to combine representationalism and phenomenal externalism with a theory of introspection are not satisfactory, I do not want to exclude the possibility that some other attempt would be.

## Summary (in Swedish)

Under de senaste åren har så kallade ”representationalistiska” medvetandeteorier ökat i popularitet. Sådana teorier beskriver mentala tillstånd som representationer av omvärlden. Dessa teorier hävdar även ofta att innehållet i mentala representationer (logiskt) bestäms av externa faktorer. Innehållet i min tanke på exempelvis vatten bestäms av förekomsten av den naturliga substansen vatten. Representationalister menar ibland även att den fenomenella karaktären hos våra upplevelser på motsvarande vis bestäms av yttre faktorer. Sanningsvillkoren för påståenden om våra upplevelser av färger involverar t.ex. ytegenskaper hos externa objekt. Detta innebär att oavsett hur de varseblivande subjekten är internt konstruerade så kommer samma påståenden om deras upplevelser att vara sanna, så länge dessa upplevelser representerar samma objekt. Denna tes har kallat ”fenomenell externalism” av t ex Fred Dretske och William Lycan.

Introspektion har traditionellt beskrivits som subjektets omedelbara medvetenhet om sina egna upplevelsetillstånd. Man har antagit att subjektet har en särskild, privilegerad tillgång till sina upplevelser, som innebär att hon inte kan missta sig om vare sig vad de handlar om eller hur de känns. En klassisk teori om hur introspektionen fungerar beskriver denna process som en slags inre perception. Subjektets tillgång till sina egna upplevelser liknar enligt denna teori i bestämda avseenden hennes tillgång till omvärlden. Denna modell av introspektion verkar även medföra att man ser på mentala tillstånd som ett slags objekt med egenskaper.

Denna syn på introspektion har visat sig vara svår att förena med fenomenell externalism. För det första verkar det följa av externalismen att subjektet inte har en epistemiskt säker tillgång till sina upplevelseinnehåll, eftersom dessa innehåll helt igenom bestäms av hur omvärlden är beskaffad. För det andra ter det sig svårt att analysera introspektion i termer av inre perception när de mentala tillstånden inte förefaller vara objekt med observerbara egenskaper. Om en upplevelse är en representation vars innehåll bestäms av externa faktorer så kan det inte vara stor idé att titta på upplevelsen för att få reda på vad den handlar om. Det finns ju ingenting som säger att representationen som sådan skulle ha de egenskaper som vi är medvetna om när vi har upplevelsen.

Det här problemet har några samtida filosofer gett sig på att försöka lösa. En del av dessa filosofer analyserar introspektion som en slags inre perception, trots att de antar att ”objekten” för denna observation saknar relevanta inre egenskaper. Lycan är en sådan filosof. Lycan hävdar också att förekomsten av inre perception av det här slaget kan förklara varför vissa mentala tillstånd är medvetna och andra omedvetna. Ett medvetet tillstånd är ett tillstånd som subjektet är medveten om genom inre perception. David Rosenthal försöker istället ge en trolig tolkning åt tesen att introspektion liknar *tänkande* på våra egna upplevelser och att det även är detta tänkande på våra egna upplevelser som gör dem medvetna. Peter Carruthers menar att vi behöver inte ägna en aktuell tanke åt ett mentalt tillstånd för att det ska vara medvetet, utan att det räcker att vi är disponerade att göra så. Medvetandetillstånd som är *tillgängliga* för sådana högre ordningens tankar är fenomenellt medvetna, dvs. subjektet har introspektiv tillgång till hur dessa tillstånd ”känns”.

Andra filosofer som också hävdar att upplevelser entydigt bestäms av sitt innehåll är starkt kritiska mot den här bilden av medvetandet. Både mot antagandet att det skulle finnas olika nivåer av medvetenhet och (i synnerhet) mot tesen att introspektion skulle likna perception. Dessa filosofer försöker ge andra analyser av vad introspektiv kunskap kan vara, som tar hänsyn till att mentala tillstånd själva saknar de egenskaper som vi är medvetna om när vi är i dessa tillstånd. De hävdar att subjektet kan vara medveten om *att* hon är i ett visst tillstånd som representerar något speciellt, men inte genom att vara direkt medveten *om* detta tillstånd. Fred Dretske, Michael Tye och Sydney Shoemaker försöker alla ge en sådan analys av introspektion.

Enligt min mening kan ingen av de teorier som just ytligt beskrivits ge en godtagbar beskrivning av introspektion, vilket jag försöker visa i denna avhandling. Avhandlingen består av två delar. I den första delen ger jag en generell översikt över teorier om introspektion och beskriver hur filosofer genom tiderna har sett på, i tur och ordning, introspektionens objekt, dess epistemologiska status samt hur den introspektiva processen fungerar. Här ser man exempel både på den mer traditionella synen på introspektion, där denna term står för subjektets omedelbara och suveräna tillgång till sina egna medvetandetillstånd, och på de analyser av introspektion som kommer att behandlas i avhandlingens andra del, där man oftast antar att att introspektionens objekt är mentala tillstånds representerade innehåll och där subjektets tillgång till detta objekt blir mer problematisk.

I avhandlingens andra del beskrivs först vad representationalism av det här slaget innebär samt hur de olika introspektionsteorierna internt skiljer sig från varandra. Jag skiljer mellan å ena sidan högre ordningens teorier (HO-teorier) och en-nivåteorier (OL-teorier). Den första typen av teori hävdar att introspektion är en högre ordningens medvetenhet som antingen är av perceptuell art eller liknar tänkande. OL-teorier framhåller att fenomenellt medvetande inte behöver förklaras i termer av högre ordningens medvetenhet utan istället kan förklaras genom den funktion mentala tillstånd spelar i det kognitiva systemet. Eftersom dessa teorier motsätter sig beskrivningen av introspektion som ett slags direkt medvetande om mentala tillstånd, kommer de också att begränsa introspektivt medvetande till subjektets kunskap att hon har ett viss upplevelse.

Jag går igenom och värderar olika argument som har framhållits mot var och en av dessa teorier och lägger även fram några egna. Min slutsats är att inget av försöken att kombinera representationalism med en positiv beskrivning av introspektion fungerar på ett tillfredsställande vis. De HO-teoretiska angreppssättet faller generellt på att dessa filosofer inte kan göra det begripligt hur vi är direkt medvetna om mentala tillstånds representerade innehåll. OL-teorierna misslyckas snarare med att göra reda för introspektiv kunskap överhuvudtaget. Subjektets självkunskap begränsas till att endast vara en kunskap om *att* man har en viss upplevelse. Hon kan t ex. inte enligt dessa teorier introspektivt skilja mellan olika *typer* av upplevelser, som mellan hörselperceptioner och synperceptioner.

## References

- Alston, W. P. 1989. Varieties of privileged access. In *Epistemic Justification: Essays in the theory of knowledge*. Ithaca: Cornell University Press.
- Anscombe, G. E. M. 1981. The first person. In *Metaphysics and the Philosophy of Mind: Collected Philosophical Papers Volume II*. Oxford: Blackwell.
- Armstrong, D. M. 1963. Is introspective knowledge incorrigible? *Philosophical Review* 62: 417-32.
- Armstrong, D. M. 1968. *A Materialist Theory of the Mind*. London: Routledge & Kegan Paul.
- Armstrong, D. M. 1980. *The Nature of Mind and other Essays*. Ithaca: Cornell University Press.
- Armstrong, D. M. 1987. *Smart and the Secondary Qualities*. New York: Basil Blackwell.
- Armstrong, D. M. 1993. Causes are perceived and introspected. *Behavioral and Brain Sciences*, 16, no 1.
- Armstrong, D. 1997. What is consciousness? In Block, Flanagan, Güzeldere 1997.
- Armstrong, D. & N. Malcolm 1984. *Consciousness and Causality*. Oxford: Basil Blackwell.
- Audi, R. 1993. *The Structure of Justification*. Cambridge: Cambridge University Press.
- Aydede, M. 2003. Is introspection inferential? In Gertler 2003.
- Ayer, A. J. 1963. Privacy. In *The Concept of a Person and other Essays*. New York: St. Martin's.
- Block, N. 1990. Inverted Earth. *Nous-Supplement: Philosophical Perspectives* 4: 4.
- Block, N. 1997. On a confusion about a function of consciousness. In Block, et. al. 1997.
- Block, N., Flanagan., O, & Güzeldere G. 1997. *The Nature of Consciousness: Philosophical Debates*. Cambridge: MIT Press.
- Boghossian, P. 1989. Content and self-knowledge. *Philosophical Topics* 17: 5-26.
- Boring, E. G. 1953. A history of introspection. *Psychological Bulletin* 50, No 3: 169-189.
- Brentano, F. 1966. *The True and the Evident*. O. Kraus, ed., R.M. Chisholm, I. Politzer, K.R. Fischer, transl., London: Routledge and Kegan Paul.
- Brentano, F. 1995. [1874] *Psychology from an Empirical Standpoint*. L.L. McAlister, ed., A.C. Rancurello, D.B. Terrell, L.L. McAlister, trans., London: Routledge and Kegan Paul.

- Brook, A. 2001. Kant, self-awareness and self-reference. In A. Brook & R. C. DeVidi 2001.
- Brook, A. & R. C. DeVidi, eds., 2001. *Self-reference and Self-awareness*. Amsterdam: John Benjamins Publishing Company.
- Brown, J. 1995. The incompatibility of anti-individualism and privileged access. *Analysis* 55, no 3: 149-156.
- Bruce, V. & P. Green. 1996. *Visual Perception: Physiology, psychology and ecology*. 3<sup>rd</sup> ed. East Sussex: Psychology Press.
- Burge, T. 1979. Individualism and the mental. In *Midwest Studies in Philosophy IV*. French, Uehling and Wettstein, eds., Minneapolis: University of Minnesota Press.
- Burge, T. 1988. Individualism and self-knowledge. *The Journal of Philosophy*, Vol. 85, No. 11, Eighty-Fifth Annual Meeting American Philosophical Association, Eastern Division: 649-663.
- Byrne, A. 1997. Some like it HOT: Consciousness and higher-order thoughts. *Philosophical Studies*, 86, no 2: 103-129.
- Byrne, A. 2001a. Review of *Phenomenal Consciousness* by Peter Carruthers. *Mind* 110, no 440: 1057-1062.
- Byrne, A. 2001b. Intentionalism defended. *The Philosophical Review* 110, no 2: 199 - 240.
- Byrne, A. 2004. What phenomenal consciousness is like. In Gennaro 2004.
- Castañeda, H.-N. 2001 [1966]. 'He': a study in the logic of self-consciousness. Reprinted in A. Brook & R. C. DeVidi 2001.
- Carruthers, P. 2000a. *Phenomenal Consciousness: A naturalistic theory*. Cambridge: Cambridge University Press.
- Carruthers, P. 2000b. Replies to critics: explaining subjectivity. *Psyche* 6(3). <http://psyche.cs.monash.edu.au/v6/>
- Carruthers, P. 2004. HOP over FOR, HOT theory. In R. J. Gennaro 2004.
- Chalmers, D. 1995. Facing up to the problem of consciousness. In J. Shear, ed., *Explaining Consciousness – The "Hard Problem"*. Cambridge, Mass: MIT Press.
- Churchland, P. M. 1985. Reduction, qualia and the direct introspection of brain states. *Journal of Philosophy* 82:1 8-28.
- Churchland, P. M. 1988. *Matter and Consciousness*. Cambridge, Mass: MIT Press.
- Comte, A. 1830. *Philosophie première. Cours de philosophie positive, Lecons 1 à 45*. Paris: Hermann.
- Dennett, D. 1988. Quining qualia. In *Consciousness in Contemporary Science*, A. J. Marcel and E. Bisiach (eds.) Oxford: Oxford University Press.
- Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown and Co.

- Descartes, R. 1973 [1911]. *The philosophical works of Descartes, Vol I*. E. S. Haldane & G. R. T. Ross, eds., and transl., Cambridge: Cambridge University Press 1973.
- Dretske, F. 1993. Conscious experience. *Mind* 102, no 406: 263-283
- Dretske, F. 1994. Differences that make no difference. *Philosophical Topics* 22, No. 1-2; 41-57. Reprinted in *Perception, Knowledge and Belief*.
- Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, Mass: MIT Press
- Dretske, F. 1996. Phenomenal externalism or if meanings ain't in the head, where are qualia? In E. Villanueva, ed., *Philosophical issues 7: Perception*. Atascadero, CA: Ridgeview Publishing.
- Dretske, F. 1999. The mind's awareness of itself. In *Philosophical studies, Vol 95*, No 1-2: 103-124.
- Dretske, F. 2000a. *Perception, Knowledge and Belief*. Cambridge: Cambridge University Press.
- Dretske, F. 2000b. What good is consciousness? Reprinted in Dretske 2000a.
- Dretske, F. 2003a. How do you know you are not a zombie? In Gertler 2003.
- Dretske, F. 2003b. Externalism and Self-knowledge. In Nuccetelli 2003.
- Evans, G. 1982. *The Varieties of Reference*. New York: Oxford University Press.
- Ewing, A. C. 1951. *The Fundamental Questions of Philosophy*. London: Routledge and Kegan Paul.
- Farah, M. 1990. *Visual Agnosia*. Cambridge, Mass: Bradford Books.
- Frege, G. 1977 [1918-1919]. *Logical Investigations*. P.T. Geach, ed., Oxford: Basil Blackwell.
- Geach, P. T. 1957. *Mental Acts: their content and their objects*. London: Routledge and Kegan Paul.
- Gennaro, R. J. 1996. *Consciousness and Self-consciousness: A defense of the higher-order-thought theory of consciousness*. Amsterdam and Philadelphia: John Benjamins.
- Gennaro, R. J. 2004. ed., *Higher-Order Theories of Consciousness: An Anthology*. Amsterdam: John Benjamins Publishing Company.
- Gertler, B. (ed.) 2003. *Privileged Access: Philosophical Theories of Self-Knowledge*. Hants: Ashgate Publishing Limited.
- Gibson, J. 1979. *The Ecological Approach to Visual Perception*. London: Lawrence Erlbaum Associates, Publishers.
- Graham, G. & L. Stephens 1994. Mind and mine. In *Philosophical Psychopathology*. Graham & Stevens (eds.) Cambridge, Mass: The MIT Press.
- Gregory, R. L. 1980. Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London B*, 290: 181-197.

- Güzeldere, G. 1997. Is consciousness the perception of what passes in one's own mind? In Block et. al. 1997.
- Hamilton, W. 1874. *Lectures on Metaphysics and Logic*. Boston: Gould and Lincoln.
- Hardin, C. L. 1986. *Color for Philosophers*. Indianapolis, IN: Hackett.
- Harman, G. 1990. The intrinsic quality of experience. In *Philosophical Perspectives 4: Action theory and philosophy of mind*. J. Tomberlin, ed., Atascadero, CA: Ridgeview.
- Hebb, D. 1977. To know your own mind. In *Images, Perception and Knowledge*. J. M. Nichols, ed., Dordrecht, Holland: D. Reidel publishing company.
- Heil, J. 1991. Perceptual experience. In *Dretske and his Critics*. B. McLaughlin, ed., Cambridge: Basil Blackwell.
- Helmholtz, H. von 1924-1925. *Helmholtz's Physiological Optics*. Translated from the 3<sup>rd</sup> edition (1909-1911) by J. P. Southwell, ed., Rochester, New York: Optical Society of America.
- Hill, C. 1991. *Sensations: A defense of type materialism*. Cambridge: Cambridge University Press.
- Hume, D. 1978 [1711-1776]. *A Treatise of Human Nature*. L. A. Selby- Bigge, ed., with an analytical index, Oxford: Clarendon.
- Husserl, E. 2001 [1900/1901]. *Logical Investigations*. Translated by J.N. Findlay from the second German edition of *Logische Untersuchungen*. London: Routledge & Kegan Paul.
- James, W. 1950 [1890]. *The Principles of Psychology*. New York: Dover Publications Inc.
- Kant, I. 1787. *Critique of Pure Reason*. Transl 1968 by N. K. Smith, London: Macmillan.
- Kobes, B. 1991. Sensory qualities and 'Homunctionalism': A review essay of W. G. Lycan's *Consciousness*. *Philosophical Psychology 4*: 147-158.
- Kobes, B. 1996. Mental content and hot self-knowledge. *Philosophical Topics 24*, no 1: 71-99.
- Krause, M. and G. Burghardt. 1999. Access to another mind: naturalistic theories require naturalistic data. *Psyche 5*. <http://psyche.cs.monash.edu.au/v5/>.
- Kriegel, U. 2002. Phenomenal content. *Erkenntnis 57*: 175-198.
- Kriegel, U. 2003. Consciousness as intransitive self-consciousness: two views and an argument. *Canadian Journal of Philosophy 33*, no.1: 103-132.
- Külpe, O. 1909. *Outlines of Psychology*. E.B. Titchener, transl, New York: Macmillan.
- Lashley, K. 1923. The behaviouristic interpretation of consciousness. In *Psychological Review 30*.
- Levin, M. 1985. Introspection. *Behaviorism 13*: 125-136.

- Lewis, C. I. 1929. *Mind and the World Order: Outline of a Theory of Knowledge*. New York: C. Scribner's Sons.
- Lewis, C. I. 1946. *The Analysis of Knowledge and Valuation*. LaSalle, IL: Open Court.
- Lewis, C. I. 1964. Some logical considerations concerning the mental. In G. N. A. Vesey, ed., *Body and Mind*. London: George Allen & Unwin.
- Lewy, C. (ed.) 1962. *The Commonplace Book of G. E. Moore*. London & New York.
- Loar, B. 1997. Phenomenal states. In Block et al 1997.
- Locke, J. 1924 [1690]. *An Essay concerning Human Understanding*. Oxford.
- Lycan, W. G. 1996. *Consciousness and Experience*. Cambridge, Mass: MIT Press.
- Lycan, W. G. 1998. In defense of the representational theory of qualia. (Replies to Neander, Rey, and Tye). In *Philosophical Perspectives, 12, Language, mind and ontology*. J. Tomberlin, ed., Oxford: Blackwell, pp. 479-487.
- Lycan, W. G. 1999. Dretske on the mind's awareness of itself. In *Philosophical studies* 95, No 1-2: 125-133.
- Lycan, W. G. 2001. The case for phenomenal externalism. *Nous-Supplement: Philosophical Perspectives 15*: 17-35.
- Lycan, W. G. 2003. Dretske's ways of introspecting. In Gertler 2003.
- Lycan, W. G. 2004. The superiority of HOP to HOT. In Gennaro 2004.
- Lycan, W. G. & Z. Ryder. 2003. The loneliness of the long-distance truck driver. *Analysis* 63, no. 2: 132-136.
- Lyons, W. 1986. *The Disappearance of introspection*. Cambridge, Mass: MIT Press.
- Lyons, W. 1991. Introspection – A two-level or one-level account? *New Ideas of Psychology* 9, no 1: 51-55.
- Lyvers, M. 1999. Who has subjectivity? *Psyche* 5.  
<<http://psyche.cs.monash.edu.au/v5/>>
- MacDonald, C. 1999. Shoemaker on self-knowledge and inner sense. *Philosophy and Phenomenological Research* 59, no 3: 711-738
- McDowell, J. 1994. The content of perceptual experience. *Philosophical Quarterly* 44, no 175: 190-205.
- McGinn, C. 1996. *The Character of Mind*, second edition. Oxford: Oxford University Press.
- McLaughlin, B. 1991. *Dretske and his Critics*. Cambridge: Basil Blackwell.
- Malcolm, N. 1963. *Knowledge and Certainty*. Englewood Cliffs, N J: Prentice-Hall.
- Malcolm, N. 1967. The privacy of experience. In *Epistemology: New Essays in the Theory of Knowledge*. A. Stroll, ed., New York: Harper & Row.
- Malmgren, H. 1983. *Immediate Knowledge: A study in G. E. Moore's epistemology*. Bodafors: Doxa.

- Mangan, B. 2001. Sensation's ghost: the non-sensory "fringe" of consciousness. *Psyche*, 7(18), October 2001. <http://psyche.cs.monash.edu.au/v7/psyche-7>
- Maund, B. 2002/2003. Tye on phenomenal character and color. E-symposium. <http://host.uniroma3.it/progetti/kant/field/tyesymp.htm>
- Maund, B. 2003. *Perception*. Bucks: Acumen.
- Mellor, D. H. 1977-78. Conscious belief. *Proceedings of the Aristotelian Society, New Series* 88: 87-101.
- Milner, D. & M. Goodale. 1995. *The Visual Brain in Action*. Oxford: Oxford University Press.
- Moore, G. E. 1922. The refutation of idealism. In *Philosophical Studies*. London: Routledge and Kegan Paul.
- Moore, G. E. 1929. Indirect knowledge. *Proceedings of the Aristotelian Society*, suppl. vol. 9: 19-50.
- Moore, G. E. 1953 [1873-1958]. *Some Main Problems of Philosophy*. London & New York.
- Moore, G. E. 1962. *Commonplace Book*. C. Lewy, ed., London: George Allen & Unwin Ltd.
- Nagel, T. 1974. What is it like to be a bat? *Philosophical Review* 83: 435 – 450.
- Natsoulas, T. 1999. The case for intrinsic theory: IV. An argument from how conscious<sub>4</sub> mental-occurrence instances seem. *The Journal of Mind and Behavior*, 20, 3: 257-276.
- Neander, K. 1998. The division of phenomenal labor: a problem for representationalist theories of consciousness. In *Philosophical Perspectives* 12, J. Tomberlin, ed., Oxford: Blackwell, pp. 411-434.
- Newton, N. 1988. Introspection and perception. *Topoi* 7: 25–30.
- Nuccetelli, S. (Ed.) 2003. *New Essays on Semantic Externalism and Self-Knowledge*. Cambridge, Mass: Bradford Books.
- Papineau, D. 2002. *Thinking about Consciousness*. Oxford: Clarendon Press.
- Peacocke, C. 1992. Scenarios, concepts and perceptions. In Crane, T. (Ed) 1992. *The Contents of Experience: Essays on Perception*. Cambridge: Cambridge University Press.
- Peirce, C. S. 1950. *The Philosophy of Peirce: Selected Writings*, J. Buchler, ed., London: Routledge & Kegan Paul.
- Penfield, W. 1975. *The Mystery of the Mind: A critical study of consciousness and the human brain*. Princeton, NJ.: Princeton University Press.
- Place, U. T. 1971. The infallibility of our knowledge of our beliefs. *Analysis* 31: 197-204.

- Putnam, H. 1975. The meaning of meaning. In *Language, Mind and Knowledge*, vol 2 of *Minnesota Studies in the Philosophy of Science*. K. Gunderson (ed) Minneapolis: University of Minnesota Press.
- Pöppel, et. al. 1973. Residual visual function after brain wounds involving the central visual pathways in man. *Nature* 243.
- Rey, G 1983. A reason for doubting the existence of consciousness. In Davidson, R., G. E. Schwartz, and D. Shapiro, eds., *Consciousness and Self-Regulation. Vol. 3*. New York: Plenum Press.
- Rey, G. 2000. Role, not content: comments on David Rosenthal's "Consciousness, content, and metacognitive judgments". *Consciousness and Cognition* 9, no 2: 147-328.
- Rivlin, R. & K. Gravelle, 1984. *Deciphering the Senses: The expanding world of human perception*. New York: Simon and Schuster.
- Robinson, W. 1999. A theory of phenomenal consciousness? *Psyche* 5. <http://psyche.cs.monash.edu.au/v5/>
- Rolls, E. T. 1998. *The Brain and Emotion*. Oxford: Clarendon Press.
- Rosenthal, D. 1986. Two concepts of consciousness. *Philosophical Studies*, 49: 329-259.
- Rosenthal, D. 1993. Thinking that one thinks. In *Consciousness: Psychological and Philosophical Essays*. M. Davies, ed., Cambridge: Blackwell.
- Rosenthal, D. 1997. A theory of consciousness. In Block et. al 1997.
- Rosenthal, D. 1999. The colors and shapes of visual experiences. In D. Fisette, ed., *Consciousness and Intentionality*. Dordrecht: Kluwer.
- Rosenthal, D. 2000. Meta-cognition and higher-order thoughts. In *Consciousness and Cognition*, 9, no 2: 231-242.
- Rosenthal, D. 2002a. Consciousness and higher-order thought. *Macmillan Encyclopedia of Cognitive Science*, Macmillan Publishers Ltd.
- Rosenthal, D. 2002b. How many kinds of consciousness? *Consciousness and Cognition*, 11, no 4: 653-665.
- Rosenthal, D. 2004. Varieties of higher-order theory. In Gennaro, 2004.
- Russell, B. 1951. *The Analysis of Mind*. London: George Allen & Unwin Ltd.
- Ryle, G. 1990 [1949]. *The Concept of Mind*. London: Penguin Books.
- Saidel, E. 1999. Consciousness without awareness. *Psyche* 5. <http://psyche.cs.monash.edu.au/v5/>
- Sellars, W. 1956. Empiricism and the philosophy of mind. In H. Feigl & M. Scriven, eds., *Minnesota Studies in the Philosophy of Science* 1.
- Sergent, J. & M. Poncet. 1990. From covert to overt recognition of faces in a prosopagnosic patient. *Brain* 113: 989-1004.

- Shoemaker, S. 1963. *Self-Knowledge and Self-Identity*. Ithaca: Cornell University Press.
- Shoemaker, S. 1968. Self-reference and self-awareness. *The Journal of Philosophy*, 65, no 19: 555-567.
- Shoemaker, S. 1994. Phenomenal character. *Noûs* 28, no. 1: 21-38.
- Shoemaker, S. 1996. *The first-person perspective and other essays*. New York: Cambridge University Press.
- Shoemaker, S. 1996b. Introspection and the self. In Shoemaker 1996.
- Shoemaker, S. 1996c. The Royce lectures: Self-knowledge and “inner sense”. In Shoemaker 1996.
- Shoemaker, S. 1996d. On knowing one’s own mind. In Shoemaker 1996.
- Shoemaker, S. 1999. Reply to Cynthia Macdonald. *Philosophy and Phenomenological Research* 59, no 3: 739-45.
- Shoemaker, S. 2000. Phenomenal character revisited. *Philosophy and Phenomenological Research*, 60, no 2: 465-467.
- Shoemaker, S. 2002. Introspection and phenomenal character. In *Philosophy of Mind*, D. Chalmers, ed., Oxford: Oxford University Press.
- Siewert, C. 1998. *The Significance of Consciousness*. Princeton: Princeton University Press.
- Skinner, B. F. 1965. *Science and Human Behavior*. Collier Macmillan – Free Press.
- Smart, N. 1964. *Doctrine and Argument in Indian Philosophy*. London: George Allen and Unwin.
- Strawson, G. 1994. *Mental Reality*. Cambridge, Mass: MIT Press.
- Strawson, P. 1959. *Individuals*. London: Methuen.
- Stumpf, C. 1906. *Erscheinungen und psychische Funktionen*. *Abh. Preuss. Akad. Wiss. Berlin (philos.-hist. Kl.)* No. 4.
- Sturgeon, S. 2000. *Matters of Mind: Consciousness, reason and nature*. London: Routledge & Kegan Paul.
- Ten Elshoff, G. 2005. *Introspection Vindicated*. Hampshire: Ashgate Publishing Company.
- Titchener, E. B. 1998 [1896]. *An Outline of Psychology*. New York: Macmillan.
- Titchener, E. B. 1912. Prolegomena to a study of introspection. *American Journal of Psychology* 23: 427-448.
- Titchener, E. B. 1912. The schema of introspection. *American Journal of Psychology* 23: 485-508.
- Titchener, E. B. 1964 [1909]. The problem of meaning. Reprinted in Mandler & Mandler, eds., *Thinking: From Association to Gestalt*. New York: John Wiley & Sons, Inc. (From Titchener, 1909, *Lectures on the experimental psychology of the thought-processes*.)

- Thomasson, A. 2000. After Brentano: A one-level theory of consciousness. *European Journal of Philosophy* 8, no 2: 190-209.
- Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge, Mass: MIT Press.
- Tye, M. 1998. Response to Discussants. *Philosophy and Phenomenological Research*, 58, no 3: 679-687.
- Tye, M. 2000a. Shoemaker's The first-person perspective and other essays. *Philosophy and Phenomenological Research* 60, no 2: 461-464.
- Tye, M. 2000b. *Consciousness, Color, and Content*. Cambridge, Mass: MIT Press.
- Tye, M. 2002. Representationalism and the transparency of experience. *Nous* 36, no. 1: 137-151.
- Tye, M. 2003. A theory of phenomenal concepts. In *Minds and Persons: Royal Institute of Philosophy Supplement: 53*, O'Hear, A. ed., Cambridge: Cambridge University Press.
- Tye, M. (forthcoming, 2006). In defense of representationalism: reply to commentaries. In *Pain: New Essays on its Nature and the Methodology of its Study*. M Aydede, ed., Cambridge, Mass: Bradford Books.
- Van Gulick, R. 2001. Inward and upward: reflection, introspection and self-awareness. *Philosophical Topics* 28: *Introspection*: 275-305.
- Van Gulick, R. 2004. Higher-order global states (HOGS): an alternative higher-order model of consciousness. In Gennaro 2004.
- Watson, J. B. 1913. Psychology as the behaviorist views it. *Psychological Review* 20.
- Watson, J. B. & W. MacDougall. 1929. *The Battle of Behaviorism*. New York: Norton.
- Wittgenstein, L. 1969 [1933-4]. *The Blue and Brown Books*. 2<sup>nd</sup> ed. Oxford: Blackwell.
- Wittgenstein, L. 1989 [1953]. *Philosophical Investigations*. Oxford: Blackwell
- Wright, C. 1998. Self-Knowledge: The Wittgensteinian legacy. In *Knowing our Own Minds*, C. Wright, B. Smith and C. Macdonald, eds., Oxford: Clarendon Press.
- Wundt, W. 1999 [1897]. *Outlines of Psychology*, (reprinted from W. Wundt, *Outlines of Psychology*, 1897), in R. H. Wosniak, ed., *Classics in Psychology, 1855 – 1914: A collection of keywords*. Bristol, Tokyo.
- Zahavi, D. 2004. Back to Brentano? *Journal of Consciousness Studies* 11, no. 10-11: 66-87.

