

# It's All in the Brain

A Theory of the Qualities of Perception

Jesper Östman



Department of Historical, Philosophical and Religious Studies  
Umeå 2013



**Department of Historical, Philosophical and Religious Studies**

Umeå University, 901 87 Umeå  
[www.idesam.umu.se](http://www.idesam.umu.se)

ISSN 1650-1748  
ISBN 978-51-7459-679-3

# It's All in the Brain

## A Theory of the Qualities of Perception

**Jesper Östman**

### Akademisk avhandling

som med vederbörligt tillstånd av Rektor vid Umeå universitet för  
avläggande av filosofie doktorsexamen framläggs till offentligt försvar i  
Hörsal F, Humanisthuset  
måndagen den 10 juni, kl. 13:15.  
Avhandlingen kommer att försvaras på engelska.

Fakultetsopponent: Professor, Howard Robinson,  
Department of Philosophy, Central European University, Budapest,  
Hungary.



**Institutionen för idé- och samhällsstudier**  
Umeå universitet  
Umeå 2013

**Organization**

Umeå University  
Department of Historical,  
Philosophical and  
Religious Studies

**Document type**

Doctoral thesis

**Date of publication**

20 May 2013

**Author**

Jesper Östman

**Title**

It's All in the Brain: A theory of the Qualities of Perception

**Abstract**

This dissertation concerns the location and nature of phenomenal qualities. Arguably, these qualities naively seem to belong to perceived external objects. However, we also seem to experience phenomenal qualities in hallucinations, and in hallucinations we do not perceive any external objects. I present and argue for a theory of the phenomenal qualities, "brain theory", which claims that all phenomenal qualities we experience are physical properties instantiated in the brain, regardless of whether they are experienced in veridical perceptions or in hallucinations.

I begin by more carefully identifying the phenomenal qualities, discussing how they are related to "qualia" and "phenomenal character". Then I present brain theory, and investigate its implications for the perceptual relations we stand in to external objects, noting that it is mostly neutral. I also compare brain theory to a similar theory of perception advocated by Bertrand Russell. Next, I provide an overview over the competing theories of phenomenal qualities, and relate them to theories of perception, such as representationalism, qualia theory, sense data theory and disjunctivism.

The majority of my argumentation for brain theory focuses on arguing that the phenomenal qualities are instantiated in the brain, rather than on arguing that they are physical properties. Instead, I largely assume physicalism. However, even independently of the physicalism assumption, I show that we have reason to believe that phenomenal qualities are experienced in hallucinations, and that qualities experienced in hallucinations are instantiated in internal objects, such as our brains or sense data. In the first step towards this conclusion I argue that theories which deny that phenomenal qualities are experienced in hallucinations face serious problems. In the next step I argue that theories which deny that phenomenal qualities experienced in hallucinations are instantiated in internal objects face serious problems. Finally, an important part of the argumentation is my replies to objections against brain theory, including common sense objections and the "observation objection". From these conclusions, together with the physicalism assumption, I infer that we have reason to believe that brain theory is true about hallucinations. On this basis, I then argue, through a generalizing argument, that the same is the case for veridical perceptions.

**Keywords**

Philosophy of perception, philosophy of consciousness, hallucination, phenomenal qualities, qualia, sense data, physicalism, Russell, disjunctivism, representationalism, brain theory, color, the brain.

**Language**

English

**ISBN**

978-91-7459-679-3

**ISSN**

1650-1741

**Number of pages**

iv, 168

# It's All in the Brain

## A Theory of the Qualities of Perception

**Jesper Östman**



**Department of Historical, Philosophical and Religious  
Studies**  
Umeå 2013

# Umeå Studies in Philosophy 11

© Jesper Östman 2013

Series editors: Sten Lindström and Pär Sundström  
Department of Historical, Philosophical and Religious Studies  
Umeå University  
SE-901 87 Umeå, Sweden

This work is protected by the Swedish Copyright Legislation (Act 1960:729)  
ISBN: 978-91-7459-679-3  
ISSN: 1650-1748

Cover image is based on a drawing by Chippolito.

Elektronisk version tillgänglig på <http://umu.diva-portal.org/>

Printed in Sweden by Print & Media, Umeå University, Umeå

Distributor: Department of Historical, Philosophical and Religious

Studies, Umeå University, SE-901 87 Umeå, Sweden. Umeå, Sweden 2013

*To mother, father and grandmother*



# Table of Contents

<b>Table of Contents</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>1. Introduction</b>	<b>1</b>
<b>2. The phenomenal qualities</b>	<b>7</b>
2.0 Introduction	7
2.1 Phenomenal qualities	7
2.2 Other ways of speaking about the phenomenal qualities	10
<b>3. Brain theory</b>	<b>15</b>
3.0 Introduction	15
3.1 Stating brain theory	15
3.2 Brain theory and spatial qualities	17
3.3 Brain theory and perceptual relations	24
3.4 Similarities to Russell	30
<b>4. Theories of the phenomenal qualities</b>	<b>35</b>
4.0 Introduction	35
4.1 Three questions about phenomenal qualities	35
4.2 II theories	38
4.3 EI theories	43
4.4 EX theories	46
4.5 EN theories	49
4.6 NN and XX theories	57
<b>5. Preliminaries to the argumentation</b>	<b>61</b>
5.0 Introduction	61
5.1 How I argue against the alternatives to brain theory	61
5.2 Definitions and theses	62
<b>6. The experience thesis</b>	<b>67</b>
6.0 Introduction	67
6.1 The knowledge-acquisition argument	67
6.2 The value argument	69
6.3 The first premise	71
6.4 The second premise	74
6.5 The no value objection	76
6.6 The disassociation objection	77
6.7 The final inference	79
<b>7. The instantiation thesis</b>	<b>83</b>
7.0 Introduction	83
7.1 Property-ism	83
7.2 Platonism	84
7.3 Aristotelianism	92
7.4 Nominalism	97
7.5 Trope theory	99

7.6 Nothingism and Meinongianism	101
<b>8. Physicalism</b>	<b>107</b>
8.0 Introduction	107
8.1 Type physicalism	107
8.2 The arguments for physicalism	107
8.3 The anti-physicalist arguments	111
<b>9. Generalizing to the veridical case</b>	<b>115</b>
9.0 Introduction	115
9.1 The first step	116
9.2 The identity argument	117
9.3 The third step	118
9.4 The causal generalizing argument	120
<b>10. Objections and replies</b>	<b>125</b>
10.0 Introduction	125
10.1 The observation objection	125
10.2 The common sense objection	130
10.3 The extension objection	133
10.4 The epistemic objection	137
10.5 The phenomenal intransitivity argument	140
<b>11. Brief summary of the results</b>	<b>147</b>
<b>Acknowledgements</b>	<b>149</b>
<b>References</b>	<b>151</b>

# Abstract

This dissertation concerns the location and nature of phenomenal qualities. Arguably, these qualities naively seem to belong to perceived external objects. However, we also seem to experience phenomenal qualities in hallucinations, and in hallucinations we do not perceive any external objects. I present and argue for a theory of the phenomenal qualities, "brain theory", which claims that all phenomenal qualities we experience are physical properties instantiated in the brain, regardless of whether they are experienced in veridical perceptions or in hallucinations.

I begin by more carefully identifying the phenomenal qualities, discussing how they are related to "qualia" and "phenomenal character". Then I present brain theory, and investigate its implications for the perceptual relations we stand in to external objects, noting that it is mostly neutral. I also compare brain theory to a similar theory of perception advocated by Bertrand Russell. Next, I provide an overview over the competing theories of phenomenal qualities, and relate them to theories of perception, such as representationalism, qualia theory, sense data theory and disjunctivism.

The majority of my argumentation for brain theory focuses on arguing that the phenomenal qualities are instantiated in the brain, rather than on arguing that they are physical properties. Instead, I largely assume physicalism. However, even independently of the physicalism assumption, I show that we have reason to believe that phenomenal qualities are experienced in hallucinations, and that qualities experienced in hallucinations are instantiated in internal objects, such as our brains or sense data. In the first step towards this conclusion I argue that theories which deny that phenomenal qualities are experienced in hallucinations face serious problems. In the next step I argue that theories which deny that phenomenal qualities experienced in hallucinations are instantiated in internal objects face serious problems. Finally, an important part of the argumentation is my replies to objections against brain theory, including common sense objections and the "observation objection".

From these conclusions, together with the physicalism assumption, I infer that we have reason to believe that brain theory is true about hallucinations. On this basis, I then argue, through a generalizing argument, that the same is the case for veridical perceptions.

**Keywords:** Philosophy of perception, philosophy of consciousness, hallucination, phenomenal qualities, qualia, sense data, physicalism, Russell, disjunctivism, representationalism, brain theory, color, brain

# 1. Introduction

*"The mind ... perceives nothing but its own ideas."*

- John Locke<sup>1</sup>

We experience certain salient qualities.<sup>2</sup> These qualities seem to me to exhaust my sensory fields: to completely fill up my visual field, my auditory field, my olfactory field and so forth.<sup>3</sup> Naively, I take them to be colors, shapes, tastes, sounds and other such properties which we experience in perception. When thinking about these qualities, I can easily focus on the ones I am currently experiencing. It also seems natural for me to say that I am aware of them in a particularly intimate and immediate way; intimate in the sense that nothing else seems to mediate my awareness of them. I will call these qualities *the phenomenal qualities*.

These qualities are very important, at least to me.<sup>4</sup> Had I not experienced phenomenal qualities it seems like a condition necessary for my experiences having any value at all, in themselves, would be unfulfilled. My life, in itself, would be meaningless.

It is perfectly clear in my mind which set of qualities I am thinking about.<sup>5</sup> However, it is not easy for me to use language to define these qualities or even to give an unambiguous definite description that points out them and nothing else. One reason for this is that most expressions I can use for such a description can be, and have been, interpreted in various different ways.

Arguably, the phenomenal qualities we experience naively seem to belong to *external* objects, that is, objects outside our own brain and mind, such as tables, trees and sandwiches.<sup>6</sup> But do they? Historically, at least since the

---

<sup>1</sup> Locke 1689, 4.4.3.

<sup>2</sup> I do not deny that we experience things other than these qualities, but these qualities are at least among the things we experience.

<sup>3</sup> At least when I have experiences in the perceptual modalities in question, for example, my smell field might be empty some of the time. The notion "sensory field" is here used in a purely non-committal way. I do not presuppose the existence of any mental objects that are fields. As we shall see below in 2.1 in addition to sensory experiences, these qualities might be involved also in non-sensory experiences.

<sup>4</sup> This value of phenomenal qualities and experiences of phenomenal qualities will be discussed in Chapters 6 and 7.

<sup>5</sup> At least it is perfectly clear which the paradigmatic cases of these qualities are. Examples of such paradigmatic cases include the qualities associated with clear colors or sharp pains. I take it that these paradigmatic qualities are part of a larger set of qualities. There might be some boundary cases for which it is less clear whether they should be included in the set.

<sup>6</sup> This is clearest for the case of color and perhaps for taste. It is a bit less clear for sounds; they generally seem external to the body but it is not obvious whether they seem to be properties of objects or objects in their own right.

time of Galileo, scientists and philosophers have claimed that the phenomenal qualities really belong to the brain, the mind or to some non-physical objects. As Galileo himself put it:

I think that tastes, odors, colors, and so on are no more than mere names so far as the object in which we place them is concerned, and that they reside only in the consciousness. Hence, if the living creatures were removed, all these qualities would be wiped away and annihilated (quoted in Drake 1957, 274).

Some contemporary scientists are of a similar opinion; if colors are anywhere they are in the brain:<sup>7</sup>

[W]e know from psychophysical and neurophysiological investigations that color is created somewhere in the brain, although the exact location of this process is still unknown, and we even have no idea what entities the sensations called color are . . . In short, colors appear only at a first naïve glance to be located in objects (Backhaus and Menzel 1992, 28).

Note here that when Galileo and Backhaus and Menzel speak about the phenomenal color qualities, the qualities we naively take to be colors, they speak of them as *colors*. One might argue that even if the qualities we naively take to be colors do not belong to external objects, a statement such as “this tomato is red” might still be true, in virtue of some property had by the external tomato being referred to in the statement. Perhaps this property is the tomato’s disposition to cause experiences of phenomenal red qualities in typical human perceivers, or else, it is some other property. This is why I use the term ‘phenomenal color’ rather than the term ‘color’ to speak about phenomenal color qualities. This issue will be sorted out more thoroughly in Chapter 2.

Today however, it is less common for philosophers to claim that the phenomenal qualities belong to the mind, the brain or some non-physical objects. Instead several philosophers, such as Michael Tye and John Campbell in the two quotes below, clearly state that these qualities belong to external objects:

[P]henomenology ain’t in the head. This is why you cannot find any Technicolor qualia, any raw feels, by peering around inside the brain ... They simply are not in there. To discover what it’s like, you need to look outside the head to what the

---

<sup>7</sup> Byrne and Hilbert (2003) think this might be the popular opinion, at least among color scientists. For more statements from scientists claiming that color is in the mind see Zeki (1983, 746) and Palmer (1999, 95-97). It seems reasonable to assume that those who believe this about color qualities also believe it about other types of phenomenal qualities.

brain states represent. Phenomenology is, in that way, externally based ... systems that are internally physically identical do not have to be phenomenally identical (Tye 1995, 151).

[The] phenomenal character of your experience, as you look around the room, is constituted by the actual layout of the room itself: which particular objects are there, their intrinsic properties, such as color and shape, and how they are arranged in relation to one another and to you (Campbell 2002, 116).

Discussions about consciousness are often framed in terms like “phenomenal character” or “qualia,” as in the quotes above. *Phenomenal character* refers to a special character of subjective experiences, what it is like subjectively to have these experiences. The term ‘qualia’ has been used in different senses, but the sense relevant here is as the qualities that constitute phenomenal character.<sup>8</sup> I believe that ‘qualia’, when used in this sense, typically refers to the phenomenal qualities, an issue that will be further discussed in 2.2.

In this dissertation I defend a theory of the phenomenal qualities that holds that they are identical to physical properties instantiated in the brain, rather than properties of external objects. I call this theory “brain theory”. My main aim is to show that we ought to raise our credence in this theory, by providing some reasons for believing it.<sup>9</sup>

Why investigate whether brain theory can be defended? The question of the location and nature of the phenomenal qualities is interesting and important in itself, since these qualities are so intimate to us and pervade our every experience. Furthermore, the question might also have crucial implications for whether the mysteries of the qualities can be solved. One reason for this is that recently, several philosophers have argued both that the qualities are external, and that this solves or ameliorates the problems of identifying them with physical properties.<sup>10</sup> But such a conceivable solution is ruled out if it turns out that the qualities rather belong to the brain. Also, some have claimed that, because phenomenal qualities belong to external objects, to understand these qualities we should look to the properties of external objects rather than to properties of the mind. But if brain theory is true, that is wrong.

---

<sup>8</sup> See for example Tye (2009) on the uses of these two terms.

<sup>9</sup> The claim that the phenomenal qualities are *physical* properties makes brain theory different from classical theories that deny that the phenomenal qualities belong to external objects, since the classical theories typically claim that the properties are non-physical. For an example of such a classical theory, see Jackson (1977).

<sup>10</sup> Loar (2003) and Byrne (2006) discuss this line of thinking critically. See Fish (2009, ch.3) for an example of a philosopher arguing that if the phenomenal qualities are external this might ameliorate the problem of identifying them with physical properties.

Another reason to investigate brain theory is that it has received relatively little recent attention, while at the same time, it seems to me, we have some reasons to believe it.<sup>11</sup> Although there are philosophers who take the phenomenal qualities to be physical properties of the brain, to the best of my knowledge, there are no contemporary publications that work out the full implications of that claim and give a sustained defense of it as a theory of the phenomenal qualities. Another reason to undertake this investigation is that, as far as I can see, the type of argumentation I use here is distinct from the forms of argumentation that have been used to defend similar theories.<sup>12</sup> Finally, a factor that further increases the importance of investigating whether brain theory can be developed and defended in the light of the scant attention it has received from philosophers is the fact that similar views of perception seem to be common among scientists, including perception scientists.

Two theses are central in my argumentation for brain theory. The first is type physicalism, which says that every actually instantiated property is a physical property. The majority of my argumentation for brain theory will focus on arguing that the phenomenal qualities are instantiated in the brain, rather than on arguing that they are physical properties. The reason for this is that giving a full treatment of either the arguments for, or against, physicalism is beyond the scope of the dissertation. Since I will not be able to provide a full defense of physicalism, the type physicalism thesis is largely assumed, rather than defended.

The second central thesis is the instantiation thesis. It says that all phenomenal qualities we experience in hallucinations are instantiated in existing “internal” objects. Internal objects are objects internal to our mind, such as brains or sense data. This thesis is entailed by brain theory, but does not entail brain theory. I will defend the instantiation thesis by arguing that theories of the phenomenal qualities which deny the instantiation thesis face significant problems, whereas at least one theory which accepts it, brain theory, does not face significant problems. Where my argumentation for brain theory rests on the type physicalism assumption, the instantiation thesis is an important result that is independent of physicalism.

---

<sup>11</sup> As far as I know, the only well known and clear earlier example of a brain theory is Russell (1927a). The biological realism of Revonsuo (2000) is another similar theory, although it is committed to the more specific claim that the qualities are *biological* physical properties of the brain. Furthermore, if it is true that philosophers who use the term “qualia” are typically speaking about phenomenal qualities then it follows, as will be discussed in Chapter 4, that some versions of the theories of perception “representationalism”, “Adverbialism” and “qualia theory” are also versions of brain theory.

<sup>12</sup> The general structure of the argumentation, which consists of first focusing on hallucination and then generalizing the conclusion to all types of perception, is old. However, most of the arguments I use are new. Throughout this dissertation, when I use old arguments, I point this out in footnotes.

Below I will briefly summarize the contents of the different chapters in my dissertation:

In Chapter 2 I introduce the term “phenomenal quality” more formally, and explain what the phenomenal qualities are. Then, I discuss how this term is related to terms like “phenomenal character” and “qualia.”

In Chapter 3 I first present brain theory. Then, I note that brain theory has some surprising implications for spatial qualities and spatial relations, while explaining what the theory says in more detail. After this, I look at what brain theory implies for the perceptual relations we stand in to external objects, noting that it is mostly neutral. Finally, I give a brief overview of similarities and differences with an older theory of perception that might seem similar to brain theory, a theory of perception advocated by Bertrand Russell.

In Chapter 4 I distinguish three different questions about the phenomenal qualities, the “kind question”, the “location question” and the “experiencing question”. Then, I present a taxonomy of different types of theories about the phenomenal qualities, and look closer at what the types of theories in it say.

In Chapter 5 I present the structure of my argumentation more carefully, noting how theories about the phenomenal qualities that are not compatible with brain theory are ruled out. Here I also introduce and explain theses and definitions which are central for the rest of the argumentation, and argue for one of these theses, the hallucination\* thesis.

In Chapter 6 I argue that theories face serious problems if they deny the experience thesis, a thesis which claims that we experience phenomenal qualities in hallucinations.

In Chapter 7 I argue that theories which agree that we experience phenomenal qualities in hallucinations, but deny that these qualities are instantiated, face serious problems. This is one part of the argumentation for the instantiation thesis.

In Chapter 8 I briefly overview the motivations for physicalism and the objections against it, striking a blow for physicalism by presenting replies to two main objections.

In Chapter 9, I argue for the conditional claim that brain theory can be reasonably inferred from the instantiation thesis, the experience thesis, type physicalism and the hallucination\* thesis, taken together.

In Chapter 10 I respond to various objections against brain theory (including the instantiation thesis). These objections include the “observation objection,” various common sense objections, the “extension objection,” the “epistemic objection” and the “phenomenal intransitivity argument.” This concludes the argumentation for the instantiation thesis and for brain theory.

In Chapter 11 I briefly summarize the main results of my argumentation.

## 2. The phenomenal qualities

*No man, who reflects, ever doubted, that the existences, which we consider, when we say this house, and that tree, are nothing but perceptions in the mind.*

-David Hume<sup>13</sup>

### 2.0 Introduction

In this chapter I introduce the term “phenomenal quality” more formally, and explain what the phenomenal qualities are. Then, I discuss how this term is related to terms like “phenomenal character” and “qualia.”

### 2.1 Phenomenal qualities

As was mentioned in the introduction, it is hard to give a strict definition of the term “phenomenal quality.” However, a fact that is potentially useful for pointing out which qualities that term refers to is that actual naïve subjects take colors, shapes, sounds, tastes, smells and pains to be phenomenal qualities.<sup>14</sup> Thus, we can point out some paradigmatic phenomenal qualities using this fact, and then later, if needed, investigate whether there is anything common to these paradigm cases that can be used to extend our grasp of the concept to non-paradigm cases. Thus:

- (1) The qualities actual naïve subjects typically take to be colors, shapes, sounds, tastes, smells and pains are paradigmatic phenomenal qualities.

Below, after introducing some terminology, I will examine the different parts of (1) more closely, partly by explaining what ‘naïve subject’ means, and then discussing what non-paradigmatic phenomenal qualities there are.

---

<sup>13</sup> Hume (1758, 152).

<sup>14</sup> Since a naïve subject is naïve about perception, we cannot ask her directly what she takes to be “the phenomenal qualities”. However, we can ask a subject which used to be naïve, but no longer is naïve, what she took to be the phenomenal qualities when she was naïve. Below, I will specify the precise meaning of “naïve” as I use the term in this context.

Let us use the term *sensible property term* for the terms used to describe what naïve subjects take to be phenomenal qualities, such as ‘color’, ‘shape’, ‘sound’, ‘taste’, ‘smell’ and ‘pain’.<sup>15</sup> Note that, since shape is included here, the properties denoted by these terms are distinct from the properties traditionally taken to be the “secondary qualities.” By (1) there is a phenomenal quality type corresponding to each sensible property term. I will use the convention that we name this phenomenal quality type by adding the adjective ‘phenomenal’ to the sensible property term. Thus, for example ‘color’ corresponds to phenomenal color, ‘redness’ corresponds to phenomenal redness, ‘pitch’ corresponds to phenomenal pitch and so on.

Note that this choice of terminology implies neither that the sensible property terms cannot refer to phenomenal qualities nor that they must refer to them. For example, although we naively take phenomenal red and phenomenal shape to be the referents of ‘color’ and ‘shape’ it is far from clear that these sensible property terms refer to phenomenal qualities. This is the reason we cannot just say that color, shape and so on *are* paradigmatic phenomenal qualities but must take the route through how things seem to naïve subjects.

For an example of how sensory property terms might refer to something other than phenomenal qualities, let us look at color terms. Someone might consistently believe that brain theory is true, so that phenomenal qualities are instantiated in the brain, while at the same time holding a theory of color terms, according to which sensible property terms refer to reflectance properties of external objects, as in the theory of Byrne and Hilbert (2003).<sup>16</sup> In this case, color terms would refer to reflectance properties rather than phenomenal colors.

On the other hand, pain provides an example where it intuitively might seem more reasonable that the sensible property term ‘pain’ refers to a phenomenal property, namely phenomenal pain, even if brain theory is true. This is because it is less clear that we take pain to be a property of objects

---

<sup>15</sup> It is a substantive question how specific or generic phenomenal qualities are. Among colors and other sensible properties one often distinguishes between determinable properties like ‘red’ and determinate properties such as a specific shade of red, perhaps the one we might call “red<sub>41</sub>.” It is clear that, if there are phenomenal qualities at all, there exist determinate phenomenal qualities such as phenomenal red<sub>41</sub>. One might wonder if there also are determinable phenomenal qualities like a phenomenal redness which does not need to be an instance of any particular determinate phenomenal redness. Here we can note that when I use an expression such as “phenomenal red” I do not mean to refer to a general phenomenal quality; rather, I have not specified which particular shade of red I am discussing.

<sup>16</sup> Note that Byrne and Hilbert themselves do not hold this combined view. They claim that color terms refer to reflectance properties, but their theory also seems to say that phenomenal qualities are reflectance properties of external objects.

external to the brain, such as body parts.<sup>17</sup> For other cases, like sound, taste and smell it seems less clear whether their sensory property terms would refer to external objects or not, if brain theory is true. This is probably because it is not clear whether we tend to take these properties to be external.

Below I will sometimes use the term ‘sensible properties.’ When using it I just use it as a convenient abbreviation for “the properties which sensory property terms refer to.” Thus, it is important to note here that I do not mean to say that there are two separate sets of metaphysical issues here; one set of questions about the nature of the phenomenal qualities and another about the nature of a second set of qualities called the sensible properties. Rather, the metaphysical issues I am considering in this thesis relate only to the phenomenal qualities, and are questions such as whether the phenomenal qualities exist, what their nature is and where they are instantiated. When we have answers to these questions, as I see it, any remaining questions about sensible properties, that is, questions about the reference of sensible property terms, will be linguistic questions.<sup>18</sup>

Let us move on in our clarification of (1). A *naïve subject* is a subject who lacks scientific and philosophical beliefs about perception. For example, it might be plausible to assume that a typical five year old child is naïve in this sense. The description (1) implies that such a child would take phenomenal color qualities to be colors. She would say that phenomenal colors belonged to the same objects as she would say were colored. If she spoke English she would use color words such as “red” and “green” to talk about phenomenal redness and phenomenal greenness, and so forth.

The description (1) is not meant to be without exception. It concerns what naïve subjects *typically* take the sensible property terms to refer to, and so is compatible with a few naïve subjects taking something other than phenomenal qualities to be the referents. Note also, that this only concerns *actual* subjects, that is, subjects in the actual world; it does not concern subjects from some other possible world. This is because it could have been the case that the world was so different from how it is, so that even naïve subjects would have taken paradigmatic phenomenal qualities to be something distinct from colors, shapes, sound, tastes, smells and pains.

---

<sup>17</sup> See Aydede (2013) on our tendency to internalize pain, taking it to be a property of experiences. However, not everyone agrees that we typically have that view of pain. An alternative account, that holds that we perceive pain as external, in the same way as color or shape, is provided by Tye (2000).

<sup>18</sup> One might, of course, still view that the questions about the reference of the sensible property terms as substantive and philosophically interesting, even after we have a theory of the phenomenal qualities.

I believe that phenomenal qualities constitute a natural kind, that is, a kind that exists independently of human classifications.<sup>19</sup> Still, since we have had a great deal of exposure to phenomenal qualities, we should have some understanding of this kind even beyond the most paradigmatic cases. Thus, we should be able to use this understanding to extrapolate from what the most paradigmatic qualities have in common in order to point out some non-paradigmatic phenomenal qualities which are nevertheless relevantly similar to the paradigm cases so as to be members of the kind.<sup>20</sup>

What would be examples of such less paradigmatic phenomenal qualities? The paradigm cases were all cases where we naively took the phenomenal quality to be something that we can perceive. Following this, examples of other phenomenal qualities are those we naively take to be the properties perceived in other modalities, such as touch and the bodily senses.<sup>21</sup>

Then there might also be phenomenal qualities which we experience, but which are not associated with sensory modalities, in other words, non-sensory phenomenal qualities. Obviously, feelings and other experiences associated with emotions seem to include such phenomenal qualities. Other even less paradigmatic examples are the qualities included in experiences one supposedly has when imagining or when thinking, so-called cognitive or non-sensory phenomenology.<sup>22</sup>

## **2.2 Other ways of speaking about the phenomenal qualities**

In Chapter 4, below, I will discuss the alternative theories about the phenomenal qualities. A problem there is that it can be hard to provide clear examples of philosophers who defend the respective alternatives, since other philosophers seldom discuss phenomenal qualities in the same terms as I used when I pointed them out. So in this section I will consider how and

---

<sup>19</sup> On the metaphysics and semantics of natural kinds, see Bird and Tobin (2010).

<sup>20</sup> The notion of relevant similarity here is vague. However, because of the previously mentioned problems with defining phenomenal qualities it might not be possible to give a precise specification of when there is just enough similarity. Relevant dimensions of similarity are those properties of paradigmatic phenomenal qualities that were mentioned in the introduction, such as our being directly aware of them, their being important to us and our difficulty in explaining them in words. To this we might add that they should be accessible in a way that is similar to the way we access paradigmatic phenomenal qualities, such as phenomenal pain or phenomenal red.

<sup>21</sup> Perhaps we can perceive far more properties through our senses than paradigmatic perceivable properties such as color and shape. If that is the case, these might also have corresponding non-paradigmatic phenomenal qualities. For example, Siegel (2006) has argued we can perceive higher-order properties such as the property of being a pine-tree. If that is the case, then there might, for example, also be phenomenal pine-tree qualities.

<sup>22</sup> For an introduction to cognitive phenomenology, see Bayne and Montague (2011).

when we can discover the position on the phenomenal qualities of philosophers who discuss them using other terms.

Rather than speaking in terms of ‘phenomenal qualities’ or ‘what we naïvely take to be color,’ it is more common to use terms such as ‘phenomenal character’, ‘qualia’ and ‘it being like something’ having an experience. These latter three notions are interrelated. *Qualia* are often thought to be the properties which constitute *phenomenal character* and an experience with phenomenal character is defined as one which *it is like something* having.<sup>23</sup> Let us for simplicity focus on the term ‘qualia’ and how it is used. Because of the relations to the other two notions it will be easy to see what conclusions about the relation between phenomenal qualities and qualia implies for the relations between phenomenal qualities and phenomenal character or “what it is like.”

In most cases I believe it is reasonable to assume that the term ‘qualia’ is used to refer to the phenomenal qualities. After all, qualia are often supposed to have many of the features I ascribed to the phenomenal qualities in the introduction: we can be aware of them in a particularly direct and immediate way, they exhaust our sensory fields and they are very important to us.<sup>24</sup>

Unfortunately, in some cases it is not obvious whether a certain theorist refers to phenomenal qualities or to something else when she writes ‘qualia’. These are cases where she merely states that the term ‘qualia’ refers to properties of the brain and that sensible property terms refers to properties of external objects, without giving any further hints about which of these two sets of properties are the phenomenal qualities. This is problematic because such claims are compatible with two different theories of the phenomenal qualities. First, the theorist might take phenomenal qualities to be external to the brain, using sensible property terms to refer to them and use the term ‘qualia’ to refer to some distinct property, internal to the brain. Second, she might take phenomenal qualities to be internal and use the term qualia to refer to them but hold that sensible property terms refer to external properties.

One circumstance which might have contributed to the existence of this problem is that qualia are sometimes defined to be properties of

---

<sup>23</sup> This is the term ‘qualia’ used in its broad sense, as the properties which constitute phenomenal character. It is sometimes also used in a distinct more narrow sense which in addition stipulates that qualia should have some combination of the properties of being irreducible, non-representational and ineffable. See Tye (2009).

<sup>24</sup> On direct awareness of qualia, see Loar (2003) and Chalmers (1996, 189). An issue related to the last of these three points, that phenomenal qualities, or at least experiences of them, are important to us, will be developed and argued for in Chapter 5 and 6.

*experiences*.<sup>25</sup> Some philosophers take experiences to be events that happen in the brain, which in that sense are *internal* to the brain. If such a philosopher defines ‘qualia’ as properties of experience, even if she thinks the phenomenal qualities are external to the brain of the perceiver, she cannot express this fact in terms of ‘qualia’.<sup>26</sup> Thus she might choose to let the term ‘qualia’ denote something other than the phenomenal qualities and instead use sensible property terms to speak about the qualities, saying that properties such as color and shape belong to external objects while insisting that qualia and phenomenal character are internal properties, properties of the brain.

Fortunately, for the purposes of this thesis, not always being able to know what a particular philosopher says about the phenomenal qualities does not constitute a serious problem. The different positions on the phenomenal qualities I will discuss are interesting in themselves, regardless of which philosophers believe in them. Furthermore, for each position we will be able to distinguish some clear cases of actual philosophers who hold the position. These clear cases are either cases where the situation described above does not arise or cases where the theorist in question gives additional clues for where she thinks the phenomenal qualities are located. Below I will give four examples of such situations where either the situation above does not arise or where we have evidence that a theorist is talking about the phenomenal qualities.

First, consider a case in which someone says that both sensible properties and qualia are properties of external objects, and does not mention any other set of internal qualities.<sup>27</sup> Here it is clear that they take phenomenal qualities to belong to external objects. The reason is that although they do not mention phenomenal qualities explicitly, the only good candidates for being phenomenal qualities are sensible properties and qualia. Since both of these are external, it follows that phenomenal qualities must be external.

Second, if a theorist does not talk about phenomenal character at all, but just applies sensible property terms to properties of external objects, perhaps even explicitly denying that there is an additional inner set of qualities, we have some evidence that she takes the qualities to be external. In the former case where she denies an additional internal set of qualities exists, this is clear, since then there are no good internal candidates for identification with the phenomenal qualities. But even when there is no explicit denial we have

---

<sup>25</sup> See e.g. Byrne (2002).

<sup>26</sup> Some philosophers think experiences can extend outside the brain but many others still believe that experiences are internal to the brain.

<sup>27</sup> An example is Tye (2000).

some evidence since it is likely that a theorist who discusses these matters and believes that there is a second internal set will mention that fact, at least sometimes.

Third, in a case where someone claims experiences have qualia and that qualia are internal by definition, but at the same time maintains that qualia lack some or all the properties I mentioned in the introduction as belong to phenomenal qualities, we also have some evidence that she takes the phenomenal qualities to be external. For example, she might say that these qualia are not important or that we cannot become directly aware of them.

Fourth, on the other hand, someone who says that sensible property terms refer to internal properties, such as the scientists mentioned in the introduction, likely takes the qualities to be internal. The reason for this is that without any external sensory properties there are no good external candidates for identification with the phenomenal qualities.



## 3. Brain theory

### 3.0 Introduction

In Chapter 3 I first present brain theory. Then, I note that brain theory has some interesting implications for spatial qualities and spatial relations, while trying to explain what the theory says in more detail. After this, I look at what brain theory implies for the perceptual relations we stand in, noting that it is mostly neutral. Finally, I give a brief overview of similarities and differences with an older theory of perception that might seem similar to brain theory, a theory of perception advocated by Bertrand Russell.

### 3.1 Stating brain theory

Brain theory says that the phenomenal qualities we experience are physical properties instantiated in the brain. We might express it more precisely as follows:

*Brain theory* (BT): Whenever a subject S experiences a phenomenal quality Q, there is a physical property P, such that P is instantiated in her brain and  $P=Q$ .

Note that brain theory is not committed to metaphysical, physiological or semantic claims that go beyond BT. In the rest of this section, I will look at a few such further questions, and explain that brain theory is not committed to any position on them, but is instead neutral. These questions include the metaphysics of properties, the specific location in the brain of phenomenal qualities, the specific kind of physical properties the qualities are identical to, and the semantics of sensible property terms.

First, in saying that the phenomenal qualities are *properties*, brain theory does not presuppose any particular metaphysics of properties.<sup>28</sup> The description of the qualities as properties in brain theory is thus supposed to be non-committed about the nature of properties. For example, they are not

---

<sup>28</sup> For an overview of different metaphysics of properties see Chapter 7.

assumed to be abstract objects, as properties are on some metaphysical positions.<sup>29</sup>

Furthermore, the statement that phenomenal qualities are identical to properties instantiated in the brain does not specify in more detail what these qualities are properties of. To give a few examples, phenomenal qualities might be (1) properties of the brain, (2) properties of some part of the brain or (3) properties of brain states. For sake of simplicity I will describe them as “properties of the brain” for the remainder of this dissertation, except where my argumentation would treat (1)-(3) differently.<sup>30</sup>

Brain theory is also neutral on the question of what kind of physical brain properties the Ps are, (i.e. the physical properties phenomenal qualities are identical to according to brain theory). The Ps might be biological, computational or even quantum mechanical properties. In particular, this means that brain theory is compatible with theories claiming that the qualities are higher-order properties - properties of properties. An example of such a higher order property is provided by a functional property of a brain property, such as the property of realizing a certain type of information processing. Since that functional property is a property of a property, it is a higher order property. Thus, brain theory does not rule out the possibility that the qualities are multiply realized, since the qualities could be identical to such higher order functional states that could be realized by different first order physical states in different creatures. For example, assuming that future computers may possibly be sentient, a certain quality might be identical to a functional information processing property that is realized by a brain property in humans, but instead realized by a computer processor property in a sentient computer.

One reason for this neutrality is that it seems to me that the empirical science of consciousness is still in its infancy. Several foundational questions are unresolved, and one might argue, as Revonsuo (2000) does, that because of this there is no widely shared empirical research program about consciousness. For example, among those researchers who agree that we

---

<sup>29</sup> For example, on Platonism about properties they are abstract objects. The same is the case on some versions of reductionist nominalism. I discuss different metaphysical positions on properties in more detail in Chapter 7 below. However, I will not there assume the truth of any particular metaphysical position about properties, but will rather argue separately for each position, to the effect that if the position is true, brain theory can better account for the phenomenal qualities in hallucinations than the rival positions. If we had reason to believe that one particular theory of properties is true, this would likely strengthen the case for brain theory rather than the opposite, since it would render redundant arguments that claim that brain theory better accounts for phenomenal qualities in hallucinations than the rival theories

<sup>30</sup> This happens when I discuss the extension objection to brain theory, in section 10.3. There I will discuss (1)-(3) separately.

should look for consciousness in the brain, there are broad disagreements about the functional level on which we should look for a correlate of consciousness. Suggestions range all the way from the quantum level (Hameroff 1998) through the levels of receptors (Flohr 2000), neurons, large populations of neurons (Singer, 1999), and pathways (Milner and Goodale 1995) to even higher functional and cognitive levels.

Finally, note that brain theory is not committed to any particular theory of the semantics of sensible property terms. Thus brain theory is also neutral on the question whether or not the phenomenal qualities are the referents of sensible property terms. It might be the case that the sensible property terms refer to phenomenal qualities as on one of the theories of sensible qualities mentioned in 2.1. Then brain theory implies that sensible properties belong to the brain. In this case, colors, smells, sounds and so forth would literally be located in the brain. But on the other hand it might also be the case that sensible property terms refer to properties of external objects, such as reflectance properties, as the other theory of sensible qualities mentioned in 2.1 claimed. In that case, sensible properties are distinct from phenomenal qualities. In this latter scenario, colors, smells, sounds and so forth could belong to external objects.

### **3.2 Brain theory and spatial qualities**

In this section I will look at the consequences of brain theory for our naïve worldview, in particular the theory's implications for phenomenal shape qualities and phenomenal spatiality.

Brain theory can have radical implications for our naïve worldview. Naively, we take paradigmatic phenomenal qualities to be properties such as color, shape, sound, taste, smell and pain. Arguably, we also naively take at least color and shape, and perhaps also sound, taste and smell to belong to external objects.<sup>31</sup> Some have even argued that pains are, or seem to be,

---

<sup>31</sup> Note that by our conception of external (i.e. external to your brain), something is external even if it seems to belong to the body or if it is located in the body or at the surface of the body. Sounds arguably, at least naively, seem to be external. See for example Casati and Dokic (2011) describing auditory phenomenology as follows: "As we have seen, we do hear sounds both as externalized (hence auditory content is at odds with proximal views) and as distally located (hence auditory content is at odds with medial views)."

Smells and tastes are less clear-cut cases. Arguably, smells seem to be located either at objects they originate from, in the air in which we smell them, or at our smell receptors in the nose. Similarly, tastes arguably seem to be located either in the objects that taste, or in our taste receptors. Although neither smell nor taste clearly

located in parts of our bodies rather than in the mind.<sup>32</sup> Thus, our naïve picture of the world is one where the world outside us is full of phenomenal colors and shapes, among other phenomenal qualities.

However, according to brain theory, phenomenal qualities are only located in our brains. It is worth emphasizing that it is not only phenomenal color properties, such as the property we naively take to be redness, that are located in the brain. Rather, the same is also the case for phenomenal shape properties, the experienced properties we naively take to be shapes, such as what we naively take to be squareness or circularity.<sup>33</sup>

Locating phenomenal shapes inside the brain has the perhaps surprising implication that the space in which phenomenal shapes extend is distinct from the space in which physical objects extend. Let us call the space phenomenal shapes extend in *phenomenal space*, and the space physical objects extend in *physical space*.

The reason why the phenomenal space and the physical space are distinct on brain theory is the following. If the phenomenal qualities we experience, including phenomenal shape qualities, are located in our brains, then these qualities do not have the spatial properties they are experienced as having, relative to physical space. But they have these properties relative to the space we experience, phenomenal space. Thus, physical space and phenomenal space must be distinct.

---

seem to be located at objects *external to the body*, in the same way as color and shape all the above-mentioned candidate loci for tastes and smells are external to the brain, and thus external in the sense relevant here.

<sup>32</sup> E.g. Tye (1997) states that pains are represented as being located in the body:

... [a] twinge of pain represents a mild, brief case of damage. A throbbing pain represents a rapidly pulsing disorder. Aches represent regions of damage inside the body rather than on the surface. These regions are represented as having volume, as gradually beginning and ending, as increasing in severity, as slowly fading away. The volumes so represented are not represented as precise or sharply bounded. This is why aches are not felt to have precise locations, unlike pricking pains, for example. A stabbing pain is one that represents sudden damage over a particular well-defined bodily region. This region is represented as having volume (rather than being two-dimensional), as being the shape of something sharp-edged and pointed (like that of a dagger). In the case of a pricking pain, the relevant damage is represented as having a sudden beginning and ending on the surface or just below, and as covering a very tiny area. A racking pain is one that represents that the damage involves the stretching of internal body parts (e.g., muscles). (Tye 1997, 333)

See also Tye (1995), (2006). If pains are represented as being located in the body outside the brain, reasonably they will also naively seem to be located in the body outside the brain and thus seem external in our sense.

<sup>33</sup> Another quality which is located in the brain, perhaps surprisingly, is phenomenal *motion*.

For example, take a veridical perception where you experience a red rectangle as being located in front of you, looking straight at it. The field of red phenomenal qualities that constitute this experienced red rectangle has certain spatial properties, relative to the space we experience. We experience it as being extended in two dimensions, the width and height dimensions, and to have a rectangular shape in these two dimensions. Furthermore, we might experience it as being located in front of us, or in other words, as being located at another location than ourselves in a third dimension, the depth dimension.

Now, according to brain theory, relative to physical space all these qualities that constitute the experienced red rectangle are located in your brain. These qualities are not located in front of your body. Presumably, they are not even located in a part of the brain that extends in two spatial dimensions, forming a rectangle.<sup>34</sup> Therefore, these qualities do not have the same spatial properties relative to physical space, as they have relative to the space we experience, phenomenal space. Thus, we may conclude that on brain theory physical space is distinct from phenomenal space.

So, given that phenomenal space is distinct from physical space, we might wonder what phenomenal space is, and what relation it bears to physical space. As I see it, talk about phenomenal space is nothing more than a convenient way to talk about the phenomenal spatial character of our experiences, and the ways phenomenal qualities are experienced as being related to each other. Thus, in talking about phenomenal space I am not committing myself to an extra metaphysical entity, phenomenal space. Phenomenal space is not a mysterious ghostly space, which is just like physical space but exists separated from it and which is literally full of mental objects. Rather, it is convenient to use the phrase “phenomenal space”, as well as other spatial terminology, such as “shape” and “position”, when describing the experienced spatial character of the phenomenal qualities that we experience.

Despite this lack of metaphysical commitment, we should acknowledge that there is an important similarity between the two “spaces”, namely that phenomenal “space” and physical space can both usefully be described by the same mathematical relations and terminology, in particular, geometry. Thus, despite that the phenomenal “space” and the physical space are very

---

<sup>34</sup> In fact, it seems very improbable that the physical properties the phenomenal qualities are identical to are arranged in the spatial configurations we experience the phenomenal qualities as being organized in when we have experiences. The reason is that if they were arranged in the same patterns we would always see a little replica of what an experiencing subject is experiencing, as long as it is experienced in some spatial pattern, when we looked in her brain, something which empirically does not seem to be the case. When we look inside brains we do not find such replicas.

different things, they share an abstract structural similarity. This structural similarity explains why it is reasonable to call both physical space and phenomenal space “a space”, and more generally it explains why it is useful to use spatial terminology when describing the phenomenal qualities we experience.

Let us look at what brain theory says in a particular situation, carefully described. Below I will describe a hypothetical case of a perceiver, Paula, seeing an object. Using this case, I will spell out what brain theory claims about the relevant different sets of properties involved in visual perception, including colors qualities, spatial qualities, the physical properties of the brain and the physical properties of external objects.

Paula veridically perceives a red square. She sits on a chair in front of a white wall, on which a square piece of bright red paper hangs, and looks at the paper. The room is illuminated by a lamp.<sup>35</sup> This results in Paula having a conscious experience of seeing a red square on a white background. In this experience, she experiences phenomenal redness and phenomenal squareness together.

What happens in this situation, according to our current knowledge of physics and physiology is that light waves emitted from the lamp reach the paper and get reflected by the wall and by the red paper square. When hitting the paper square, due to its reflectance properties light of some frequencies, in particular from the red wavelength spectrum, have a greater tendency of being reflected than other frequencies. This contrasts with the light hitting the white wall. From that surface most frequencies are reflected, giving rise to a different profile. So light with a certain profile is reflected from the square area, and light with a different profile is reflected from the surrounding wall. Out of all these emitted and reflected light waves some reach Paula’s eye. Cones and rods, specialized neurons in her retina, register the directions and profiles of the incoming light waves, propagating this information up through the brain, in particular through her visual centers. This results in various neural events in the brain, where ultimately representations of the wall and of the red square are reconstructed from the sensory information. Some of the resulting brain properties correlate with her experience of phenomenal red, other with her experience of phenomenal squareness. Let us call these neural correlates to the experienced phenomenal qualities  $N_r$  and  $N_s$  respectively.

---

<sup>35</sup> To simplify, let’s assume she looks down on it from just above, so she does not see it at an angle. Otherwise we face the question of whether a square seen at an angle really looks square, or if it looks like a rhomb.

Three sets of properties are relevant here. First, there are the phenomenal qualities. Second, there are the shape and surface reflectance properties of the external objects that are the best external candidates for being called “colors” and “shapes.”<sup>36</sup> Let us call these “external colors” and “external shapes.” More generally, each sensible property will have a corresponding such external property.<sup>37</sup> We can call these properties collectively “the external candidates.” Third, there are neural correlates to her experiences of phenomenal qualities.

According to brain theory, the phenomenal qualities are identical to their neural correlates. Thus Nr and Ns belong to the set of properties, P, which consists of physical properties with which the phenomenal qualities are identical, according to brain theory. The only role the external shape and reflectance properties play is to be distal causes of the experience of phenomenal shape and phenomenal color qualities.

Below, I will illustrate this in a picture. To facilitate the understanding of the picture I will introduce some new terminology to get short but precise designations for the various properties. The phenomenal colors will be called “colorsQ”, where “Q” stands for quality. For example, phenomenal redness will be called “redQ.” Similarly, the phenomenal shapes will be called “shapesQ.” So, phenomenal squareness is called “squareQ.”

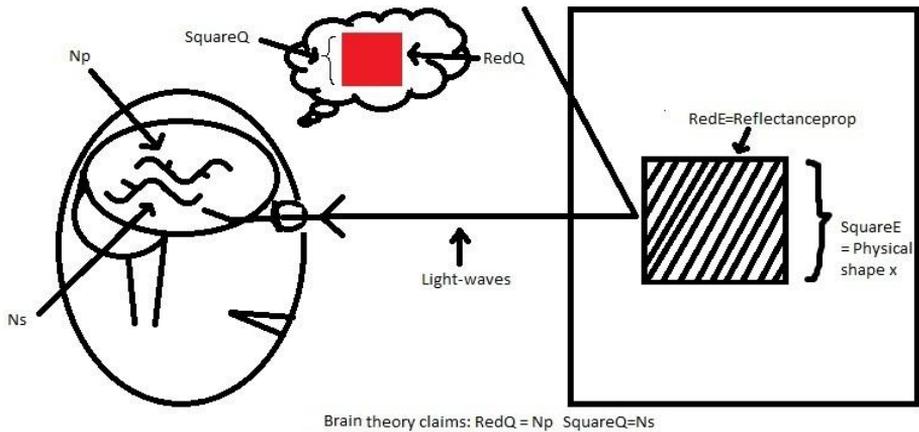
The external colors will be called “colorsE,” where “E” stands for external. In particular external redness will be called “redE.” Similarly, the external shapes will be called “shapesE.” For example external squareness will be called “squareE.”

The colors and shapes shown in thought bubbles in the pictures below depict the phenomenal shapes and colors as our perceiver Paula experiences them.

---

<sup>36</sup> According to one theory of color, color physicalism, colors are surface reflectance properties. But there are also alternative theories of color. Subjectivists take colors to be properties of experiences or of the mind. Primitivists take them to be primitive non-physical properties of external objects. Finally, dispositionalists take colors to be perceiver-dependent relational properties. On the different theories of color see Maund (2012).

<sup>37</sup> Exceptions are cases where the sensible property either lacks an external candidate, or where there are many equally good candidates. For simplicity, I will assume that each sensible property has a unique corresponding external candidate, when discussing external candidates below.



Picture 1: Paula sees a red square

Picture 1 above shows the head of Paula, our perceiver, it also shows the experienced red square, and the external red paper square that causes her experience. The picture also sketches part of the chain of causation in this perceptual event, from the light reflected from the wall to her brain. The stripes on the external square symbolize that the atoms on its surface are disposed to reflect light with a particular wavelength profile.

Her head is shown from the side, with some inner parts visible. In particular, I have shown her eye with its connection to the brain, her cortex, cerebellum and brain stem, as well as the neural structures instantiating the neural correlate properties Nr and Ns.<sup>38</sup>

In the picture, the thought-bubble shows the world as she experiences it. The red square inside it depicts the phenomenally red and square qualities she experiences, redQ and squareQ. As indicated in the picture, according to brain theory the redQ is identical with the neural property Nr and squareQ is identical with Ns.

The thought-bubble depiction is not supposed to imply that the experienced square is located outside Paula's head or outside space. Rather, it shows the

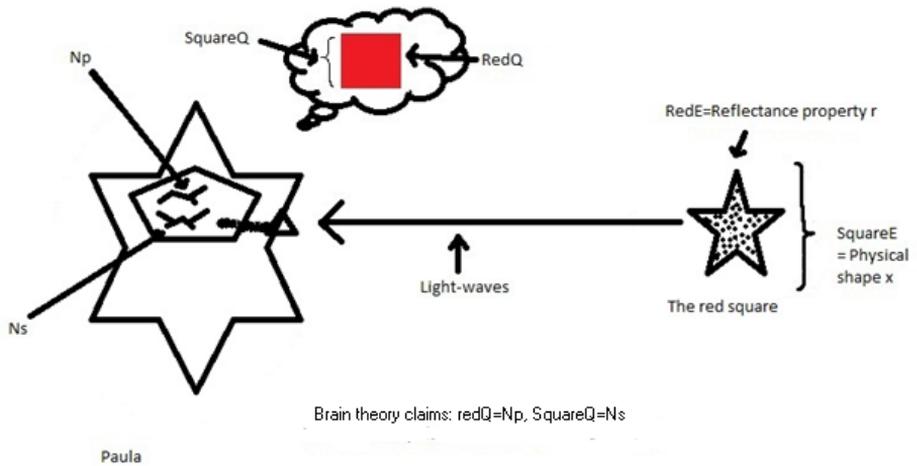
<sup>38</sup> In drawing this picture I have assumed that different structures instantiate the neural correlate properties Nr and Ns. It is also possible that they are instantiated by the same structure. This is an empirical question for the science of consciousness so brain theory is not committed to any particular theory as to whether Ns and Nr are instantiated by the same or by different structures.

phenomenal qualities as they are experienced in her phenomenal space, which exists in her brain where the qualities are located. Since according to brain theory, the experienced qualities in question are identical with Nr and Ns, they are very much located in the same locations as these. The reason for the bubble is that phenomenal space and physical space cannot both be straightforwardly depicted in the same picture. If we tried, either their phenomenal size and shape or their physical size and shape would be distorted. For example, the experienced square here does not have a square shape in the part of physical space depicted here. So if we are to show it as a square, it must be located in another space.

Furthermore, note that this picture can be misleading in one way since it portrays the physical shapes as if they were similar to phenomenal shapes, but black and white rather than colored. However, on brain theory this need not be the case. The only similarities between physical and phenomenal space that are necessary, given brain theory, are purely abstract structural relations, such as phenomenal height and phenomenal width standing in the same relation to each other as physical height and phenomenal width stand to each other.

To emphasize this fact let us look at a second picture, Picture 2, which abstracts away from some of the potentially misleading similarities in the first picture. Imagine that the physical world has a complicated geometry, where star-shaped and polygon-shaped objects interact under such laws of nature that they behave and interact in the same ways as squares and heads. In such a world we could depict the physical reality as in the picture below.

Here the shapes outside the thought bubble depict external shapes in physical space. Nr, Ns and Paula's head, brain and eye have been replaced with less familiar looking shapes from this complicated geometry.



Picture 2: Paula sees a red square

### 3.3 Brain theory and perceptual relations

One might wonder what brain theory says about questions of how and if we are perceptually related to the external world. What are the contents of perception, that is, what do perceptions represent? What are the objects of perception, that is, what do we perceive? Do we perceive external objects directly? When having veridical perceptions, is the external world as we experience it or not? We can enumerate these questions as follows:

- (1) What are the contents of perception?
- (2) What are the objects of perception?
- (3) Do we perceive external objects directly?
- (4) Is the external world as we experience it in veridical perceptions?

Let us begin with the question (1). A *content* of an experience is a condition of satisfaction that the experience has; loosely speaking a way the world can

be to make the experience “correct” or “incorrect”, much like a statement can be correct or incorrect. This condition can be satisfied or unsatisfied. If the experience is a veridical perception the content it has is satisfied. On the other hand, if the experience is a hallucination, then the content it has is not satisfied. For example, consider an experience as of a red square. The content of this experience is satisfied if the subject of the red square experience veridically perceives a red square. But if the subject is merely hallucinating, rather than veridically perceiving a red square then the content of this red square experience is not satisfied.

Brain theory is neutral on whether experiences have contents. It is also neutral on what specific conditions these contents are, if experiences have contents. We can see this if we briefly look at the different theories of why experiences have contents. What contents perceptions have is determined by different facts on different theories about the contents of perception.<sup>39</sup> On some theories, so called “internalist” theories, the contents of experiences are determined by factors that are internal to the brain, such as facts about the function of the experience in the brain. On other theories, so called externalist theories about content of perception, the contents of experiences are, at least partly, determined by factors that are external to the brain of the perceiver. These external factors can be social factors, such as under what conditions subjects in our linguistic community typically judge experiences to be veridical or falsidical. These external factors can also be causal or historical facts, such as facts about which external objects typically cause experiences. Finally, on some theories of content, experiences can have content in virtue of their phenomenal character: “phenomenal content”.<sup>40</sup> More specifically, these are representational contents of experiences that are determined by the phenomenal character of said experiences, that is, by the phenomenal qualities experienced in the experiences.

On neither theory brain theory has significant implications for what the contents of perception are. Functional, social and causal facts obtain independently of the location of the phenomenal qualities. Phenomenal content, on the other hand, might, at first sight, seem to depend on theories of the phenomenal qualities, since what phenomenal content an experience has is determined by what phenomenal qualities are experienced in it. However, brain theory does not make any claims about the character of phenomenal qualities, it only makes claims about their physicality and location, and these claims should not put any significant constraints on the

---

<sup>39</sup> On the contents of perception in general, see Siegel (2011) and Chalmers (2010).

<sup>40</sup> An example of such a theory is provided by Chalmers (2010). See also Siewert (1998).

character of the qualities.<sup>41</sup> Thus, it seems brain theory is also largely neutral on phenomenal content, and we may conclude that it does not have any significant direct implications for the contents of perceptions.

Proceeding with (2), the standard theory for what the objects of perception are is the causal theory of perception. This theory gives an analysis of perception according to which we perceive an object when having an experience, if the experience stands in an appropriate causal relation to the object. On some versions of the causal theory perceiving an object also requires that the experience bears an appropriate matching relation to the object.<sup>42</sup> Following Chalmers (2010) we might extend this analysis beyond objects, to perception of properties:

The standard view of the perceptual relation between subjects and objects holds that it is a causal relation: to perceive an object is roughly to have a perception that is appropriately caused by the object (and perhaps that has a phenomenal character that is appropriately related to the character of the object). The standard view of the perceptual relation between subjects and properties is presumably something similar: to perceive a property is roughly to have a perception whose phenomenal character is appropriately causally related to an instance of that property (and perhaps whose phenomenal character represents the instantiation of the property, or otherwise “matches” the property in some fashion).<sup>43</sup>

Below I will be adopting this theory of the objects of perception. Now, on brain theory, our experiences are distally caused by external objects in veridical perceptions. Furthermore, although its underlying nature is different, the phenomenal character of veridical perceptions is the same on brain theory as on all rival theories we are considering.<sup>44</sup> Thus, if the

---

<sup>41</sup> There is one, minor, exception. Brain theory has some limited implications for the contents of experiences, if experiences have phenomenal contents. The reason is that brain theory claims that we experience phenomenal qualities in hallucinations, whereas EX and XX theories deny this. On the latter theories hallucinations lack phenomenal qualities, which implies that they also lack phenomenal content. Thus, brain theory, along with all theories of phenomenal qualities, except EX and XX theories, entails that hallucinations have phenomenal content, *if* experiences have phenomenal contents. But because this exception is minor I do not qualify the claims above about the neutrality of brain theory regarding phenomenal content.

<sup>42</sup> On the causal theory of perception see Grice (1961) and Lewis (1988a).

<sup>43</sup> Chalmers (2010, 429).

<sup>44</sup> XX views are an exception, since they claim that veridical perceptions lack phenomenal qualities, and thus phenomenal character. However, below in Chapter 4 I am assuming that XX views are false, since they have few, if any, proponents, and arguing against them would go require extensive argumentation that goes beyond the scope of this dissertation.

Note that although they have the same character, phenomenal qualities experienced in veridical perceptions have different metaphysical locations on brain theory and on rival theories. On brain theory, phenomenal qualities are located in the brain. On most alternative theories of phenomenal qualities, qualities experienced in veridical perceptions are instead located in external objects. But this difference in location should not imply that there is a difference in the matching relation. If the phenomenal character of one experience matches a

phenomenal character matches certain external objects on rival theories, it should also match the same external objects in brain theory. So we reach the conclusion that the ordinary external objects we think we perceive in veridical perceptions actually are the objects of perception even on brain theory.

An interesting consequence of the standard theory of the objects of perception is that, on brain theory, we stand in the perception relation to external objects, but not to phenomenal qualities, since phenomenal qualities typically do not cause perceptions. Rather, when we experience phenomenal qualities we stand in a different relation to them. We might call this relation "experiencing." Note that on other theories of perception, the experiencing and perceiving relations need not be distinct relations. For example, an EX theorist might claim that the relation we stand in to external objects is identical to the relation we stand in to phenomenal qualities.<sup>45</sup> A consequence of this identity is that we cannot experience phenomenal qualities in hallucinations, since we do not stand in any relation to veridical perceptions in hallucinations.

Both the perception relation and the experiencing relation are grounded in physical processes. Brain theory's model of perception states that the perceiving process proceeds roughly according to the following description. In perception, external objects reflect patterns of light beams which are detected by the eyes. Depending on their surface reflectance properties, the objects reflect different proportions of the incident light in different parts of the visible color spectrum. This information, along with perceptual clues, and perhaps other background knowledge, is used by the brain's visual system to generate internal representations of the external objects. The phenomenal qualities and the neural correlate they are identical to, enter at the last stage, since they are plausibly properties of such representations. Finally, there is some neural causal process that connects the neural correlates to whatever in our brain it is that constitutes our subject.<sup>46</sup> This process grounds the experiencing relation, and so could aptly be called the *experiencing process*. Metaphorically speaking, we could think of this experiencing-process as mediated by an "inner eye," in contrast to the outer eye that mediates the perceiving-process in the case of seeing.

---

certain external object, then an experience with the same phenomenal character, but where the qualities that constitutes this phenomenal character has a different location, should also match that object.

<sup>45</sup> Note that disjunctivist might dispute the standard causal analysis of the perception analysis. Cf. Snowdon (1980-1981).

<sup>46</sup> This claim does not presuppose that there is a single place in the brain where all information comes together; what constitutes the subject in the brain could be some spread out network, for instance. For criticism of the idea that all the information in the brain has to come together in a "Cartesian theatre", see Dennett (1991).

Let us proceed to (3), the question whether we perceive external objects indirectly or directly, according to brain theory. There are many different ways to define “direct perception,” and the debate about which of these proposed definitions is the correct one is an extensive and complicated debate. Thus, I will merely note three different ways of defining “direct perception,” and briefly look at whether brain theory counts as direct or not, on the respective four ways:

*No-objects-directness:* Perception of an object is direct if and only if there are no mental objects.<sup>47</sup>

*Experiencing-directness:* Perception of an object is direct if and only if we stand in the experiencing relation to the object.

*In-virtue-of-directness:* Perception of an object is direct if and only if we do not perceive it in virtue of perceiving another object.<sup>48</sup>

Brain theory is only plausibly indirect on the second of these three senses of direct perception, experiencing-directness. Brain theory does not claim that there are any mental objects, such as non-physical sense data: thus it counts as a direct theory of perception on the first criterion, no-objects-directness. On the second criterion, experiencing-directness, we experience phenomenal qualities, so these are perceived directly. However, on brain theory we do not stand in the same relation to external objects, as we do to phenomenal qualities. Thus, we do not stand in the experiencing relation to them, and can at best perceive them indirectly. So brain theory is an indirect theory of perception on the experiencing-directness criterion.

Finally, on the in-virtue-of-directness criterion, brain theory is not an indirect theory of perception, as it does not posit any intermediate objects, such as non-physical sense data, in virtue of which we perceive external objects. Since we are directly aware of phenomenal qualities, it would be possible to construe a version of brain theory according to which we perceive external objects in virtue of perceiving phenomenal qualities. However, such a position is not mandatory for a brain theorist. First, as was discussed above, we do not stand in the relation perceiving to phenomenal qualities. Rather, we stand in another relation to them, which might plainly be called “experiencing.” Thus, we do not perceive external objects in virtue of *perceiving* phenomenal qualities. Second, the fact that we are directly aware of phenomenal qualities need not prohibit us from counting as directly aware

---

<sup>47</sup> For example Smith (2006, 423) endorses this criterion.

<sup>48</sup> This criterion is most famously defended by Frank Jackson, in Jackson (1977). It is criticized by Fish (2004) and Campbell (2004).

of external objects, when having veridical perceptions. The reason for this is that it might be the case that our awareness of the external objects is not mediated by awareness of phenomenal qualities, but instead proceeds independently of it, in a parallel act of awareness. In conclusion, it seems brain theory does not imply that perception is indirect. Specific versions of brain theory can be construed either as direct or indirect theories of perception.<sup>49</sup>

Finally, let us look at the question (4), whether the external world is as we experience it in veridical perception or not, if brain theory is true.<sup>50</sup> This question seems to be closely related to the question of whether the external world is as our experiences represent it as being. Thus, one natural interpretation of (4) would be as the question of whether the external world is as our veridical perceptions represent it or not. So, under this interpretation we might express (4) as the question about whether the contents of our veridical perceptions are satisfied.

As explained above, brain theory does not have any direct implications for what the representational contents of experiences are. Typical veridical perceptions arguably represent external objects and their properties, regardless of whether brain theory is true or not. Thus, on this interpretation of (4), brain theory does not have direct implications for whether the external world is as we experience it in veridical perceptions.

On the other hand one can also naturally interpret (4) as primarily being concerned with phenomenal qualities, rather than representational content. For example, we might interpret “how we experience the world” as a statement about how we naively take the world to be, rather than as a statement about how our experiences represent the world. Arguably, we naively take the phenomenal qualities to belong to external objects.

Since according to brain theory, phenomenal qualities belong to the brain, rather than to external objects, the external world is not the way we naively take it to be.<sup>51</sup> Given that this naïve view of the world is based on perception,

---

<sup>49</sup> The question whether this association with indirectness is a problem for brain theory or not will be discussed in Chapter 10.

<sup>50</sup> Note that if the world is not the way our “veridical perceptions” represent it, in a sense these perceptions are not veridical, although we take them to be veridical. For example, Chalmers (2010), who holds that experiences have multiple contents, some of which are satisfied in “veridical perceptions” and some of which are not, distinguishes between two senses of veridicality: “perfect veridicality” and “imperfect veridicality”.

<sup>51</sup> A possible exception to this is discussed in Chapter 10, the case of looking inside a brain when it has experiences.

Furthermore, one might think even if the external world does not contain phenomenal qualities it might still contain properties which are similar to phenomenal qualities. However, if we consider what kinds of

which seems reasonable, in some sense perception must present the phenomenal qualities as belonging to the external world. Thus, if brain theory is true, it follows that the external world is not as it seems in perception, in this sense.

Personally, I have some sympathy with this latter interpretation. Even now when I am no longer a naïve subject, when having perception experiences I still seem to experience the phenomenal qualities in these experiences as belonging to external objects.

### 3.4 Similarities to Russell

In this section I will compare brain theory to an earlier theory of perception which is similar to it, namely the theory of perception from Bertrand Russell's later writings. The reason it is instructive to compare brain theory with Russell's theory is that the latter both agree that phenomenal qualities are instantiated in the brain, and present similar conceptions of phenomenal space. As we shall see the two theories both have similarities and differences.

The most striking similarities concern the location of the qualities. See, in particular, Russell (1927):

As will appear when we come to the causal theory of perception, the whole of my perceptual world is, from the standpoint of physics, in my head; any two events which I experience together overlap in physical space, and all of them together, in *physical* space, occupy a volume smaller than my head, since it certainly does not include the hair, skull, teeth, etc (Russell 1927a, 145).

Here Russell claims that his “perceptual world” is located in his brain. The perceptual world includes everything he directly experiences, so reasonably, that claim implies that the phenomenal qualities we experience in perceptions are instantiated in our brains. Thus, he should agree with brain theory that whenever a subject experiences phenomenal qualities these qualities are instantiated in her brain.

---

properties brain theory says that the phenomenal qualities are we realize that this possibility is reasonably ruled out. Brain theory holds that the qualities are physical properties of brain states. But in general, brain states and their properties are very different from typical external objects such as chairs and tables.

Another similarity between Russell's theory of perception and brain theory is that Russell makes a distinction between physical space and phenomenal space (or "space in psychology," as he calls it) that is very similar to the distinction between physical and phenomenal space that brain theory makes:

At this point, it is important to be clear as to the difference between "space" in psychology and "space" in physics. There is undoubtedly a connection between the two, which it will be part of our business to make clear at a later stage. But the connection is very round-about and inferential. At the outset, it is much more useful to realize the difference between them than the connection, since much confusion of thought arises from supposing the connection to be closer than it is. In physics there is only one space, while in psychology there are several for each individual; these can, it is true, be reduced by manipulation to one for each individual, but they cannot be reduced further without introducing obscurities that it is impossible to dissipate. The space containing my visual objects has no point in common with the space containing yours, since no visual object in my world is precisely identical with one in yours. And the amalgamation of the spaces of my different senses into one space is a piece of early science, performed by the infant at about the age of three months (Russell 1927a, 143-144).

In addition to distinguishing phenomenal space and physical space, he also shares the conception of phenomenal space as not giving rise to any substantial metaphysical commitment. Rather, speaking about phenomenal space is merely a convenient way to describe our experiences:

There certainly are perceived relations between figures, and these perceived relations are part of our perceptual data in physics. Whether they are to be said to constitute a space or not, is a verbal question. Psychologists, as a rule, find it convenient to say so; but the matter is unimportant.<sup>52</sup> However, there are also important differences between Russell's theory of perception and brain theory. A main difference is that, rather than being a physicalist, Russell adhered to neutral monism, the view that reality is fundamentally neither mental nor physical but something else.<sup>53</sup> Thus, unlike brain theory, he did not endorse physicalism, the thesis everything is physical.

Furthermore, Russell's theory may have had several other controversial components that are not entailed by brain theory:

---

<sup>52</sup> Russell (1927a, 145).

<sup>53</sup> On neutral monism in general, and Russell's neutral monism in particular, see Stubenberg (2010).

I maintain an opinion which all other philosophers find shocking: namely, that people's thoughts are in their heads. The light from a star travels over intervening space and causes a disturbance in the optic nerve ending in an occurrence in the brain. What I maintain is that the occurrence in the brain is a visual sensation. I maintain, in fact, that the brain consists of thoughts-using 'thought' in its widest sense, as it is used by Descartes...What I maintain is that we can witness or observe what goes on in our heads, and that we cannot witness or observe anything else at all (Russell 1959, 18-19).

Here Russell claims that brains are, at least partly, composed of "thoughts" and that we perceive our own brains and nothing else. The former claim, if interpreted as saying that brains, or parts of them, are fully composed of thoughts, seems to imply a form of panpsychism, the view that everything is mental. But panpsychism is not entailed by brain theory. That phenomenal qualities are physical properties instantiated in the brain does not imply that all physical properties are mental. The latter claim, that we perceive nothing but our own brains, is either a statement of indirect realism, here saying that we only perceive our own brains *directly*, or it is the stronger statement that our own brains are the only things we perceive, directly or indirectly. But as we saw in the last section, although brain theory is compatible with indirect realism it does not entail that position. In particular, brain theory does not imply that we only perceive our own brains, directly or indirectly.

Another central Russellian thesis, which was related to his theory of perception, was the epistemic thesis that we cannot know anything about the intrinsic structure of the external world, except what we know from the internal world. See for example Russell (1982):

We can know all those things about physical space which a man born blind might know through other people about the space of sight; but the kind of things which a man born blind could never know about the space of sight we also cannot know about physical space. We can know the properties of relations required to preserve the correspondence with sense-data, but we cannot know the nature of the terms between which the relations hold (Russell 1982, 16).

Although I personally am sympathetic to this claim, and it might seem to naturally fit brain theory we must note that it is not strictly implied by brain theory.

In addition to the differences in the content of brain theory and Russell's theory of perception, they also have different motivations. The considerations which led Russell to believe in his theory were very different from the arguments I use to defend brain theory. For example, one of his

reasons was based on an argument from perception being causally mediated. The idea was briefly that, assuming that experience happens as the last event in a causal chain, it would be miraculous if this experience had a radically different nature from its cause. However, we could avoid this miracle if the final event was similar to its most proximate cause, which was a physical event in the brain.<sup>54</sup>

Another of Russell's reasons for his theory was based on the so called "time-gap" argument, the idea that since there is a delay between external physical events and our perception of them we cannot be perceiving them directly:

The supposition of common sense and naïve realism, that we see the actual physical object, is very hard to reconcile with the scientific view that our perception occurs somewhat later than the emission of light by the object; and this difficulty is not overcome by the fact that the time involved, like the notorious baby, is a very little one (Russell 1927a, 155).<sup>55</sup>

My argumentation on the other hand, as I sketched above in the introduction, consists of two parts. First, it uses various arguments to argue that brain theory is the best theory of the phenomenal qualities in hallucination. Then, second, it uses a generalizing argument, the "identity argument," to reach the conclusion that if brain theory is true in hallucinations it is also true in cases of veridical perception.

---

<sup>54</sup> See Russell (1927a, 400) and Russell (1927b, 111) for this reasoning. Stubenberg (2010) discusses the reasoning.

<sup>55</sup> See also Russell (1982). The argument is discussed by Huemer (2011). A common reply from opponents of the time-gap argument is that it is possible to "see into the past," in the sense of seeing an event which transmitted information to one of our senses as it was when it happened. For this response, see Cornman (1975, 49-50) and Huemer (2001, 131-5).



## 4. Theories of the phenomenal qualities

### 4.0 Introduction

In this chapter I give an overview of the main theories of phenomenal qualities. First, I will distinguish three different questions about phenomenal qualities, the “kind question”, the “location question” and the “experiencing question”. Then, I will present a taxonomy of types of theories about the latter two questions, and, in turn, look closer at the types of theories in this taxonomy. When looking at the types I will do three things. First, I describe type and any relevant subcategories within it. Second, I will note which types and subtypes are compatible with brain theory, and which provide alternatives to brain theory. In particular, I will look at what alternatives they provide for phenomenal qualities in hallucinations. Third, I will also briefly note which, if any, theories of perception that can exemplify these types of theories of phenomenal qualities.

### 4.1 Three questions about phenomenal qualities

We can distinguish between three types of questions about phenomenal qualities. First, there is the question about *whether we experience* phenomenal qualities at all. Then, there is the question of *what kind of properties* the phenomenal qualities are. Among other things, this question concerns whether the qualities are physical or non-physical properties.<sup>56</sup> Finally, there is the question of what, if anything, the phenomenal qualities *are properties of*. We can call the first *the experiencing question*, the second *the kind question* and the third *the location question*. The first question concerns our relation to the phenomenal qualities, and the other two concern the nature of the phenomenal qualities. Thus, a theory of the phenomenal qualities provides an answer to some of these questions, or to all of them.

Answers to these questions are, at least partially, independent. In principle, any view about the nature of phenomenal qualities can be combined with

---

<sup>56</sup> Other aspects of this question concern whether the phenomenal qualities belong to more specific kinds, such as functional properties or neurological properties.

either the position that we experience phenomenal qualities or that we do not experience phenomenal qualities.<sup>57</sup> An answer to the question whether the qualities are physical or not can be combined with different claims about what the qualities are properties of. For example, looking at color qualities, the view that they are non-physical can either be combined with a view holding that they are instantiated in brains, or a view holding that they are instantiated in external objects. Similarly, most answers to the question of where, if anywhere, the phenomenal qualities are instantiated can be combined with different claims about whether these qualities are physical.<sup>58</sup>

Most of this dissertation concerns the experiencing question and the location question, but not the kind question. The part of the kind question that we will consider is whether the phenomenal qualities are physical or not, which is covered in Chapter 8. The reason I do not give a full treatment of this question is, as I mentioned in the introduction, that giving a comprehensive treatment of physicalism goes beyond the scope of my dissertation. Because of this focus on the location question, the rest of this section will only cover different answers to the location and experiencing questions.

The experiencing question has two possible answers, either we experience phenomenal qualities or we do not. The location question on the other hand is more complicated. According to brain theory phenomenal qualities are instantiated in the brain. However, many theories of phenomenal qualities deny this. Some theories say that phenomenal qualities are instantiated in external objects. Other theories say that the phenomenal qualities are instantiated in non-physical objects. Finally, some theories claim that the phenomenal qualities are not instantiated at all.

Below I will give a taxonomy of different locations for phenomenal qualities. For the purposes of my argumentation for brain theory, it is most useful if we structure this taxonomy around the distinction between hallucinations and veridical perceptions.<sup>59</sup> Since some theories give different answers to the

---

<sup>57</sup> However, some combinations have never been defended, as far as I know, for example the combined view that phenomenal qualities belong to non-physical sense data *and* are never experienced.

<sup>58</sup> One exception to this independence is if the phenomenal qualities are instantiated in non-physical objects. In that case, it seems unlikely, if not impossible, that the phenomenal qualities should be physical properties, since plausibly non-physical objects cannot have physical properties. Furthermore, some kinds of properties cannot belong to any type of object. For example, neurological properties cannot belong to external objects that are not brains. Because of exceptions, such as these, the kind question and the location question are not completely independent.

<sup>59</sup> Recall, as was mentioned in the introduction, I will first argue that phenomenal qualities are experienced in hallucinations. Then in the second part of the argumentation, I will argue that we have reasons to believe that such qualities are instantiated in the brain. In the third step, I will argue that if phenomenal qualities are instantiated in the brain, then phenomenal qualities are also instantiated in the brain in veridical perceptions. Most of this argumentation is about showing that the phenomenal qualities experienced in hallucinations are

location and experiencing questions for phenomenal qualities experienced in hallucinations and phenomenal qualities experienced in veridical perceptions, it is useful if our taxonomy distinguishes theories by what they say about the qualities in hallucinations and veridical perceptions respectively. For simplicity, let us call being in a veridical perception “the good case” and being in a hallucination “the bad case.”<sup>60</sup> Philosophers sometimes talk about “illusions” as a third category of experiences, distinct from both veridical perceptions and from hallucinations. The core idea is that in an illusion some external object is perceived, but it is perceived as being different from how it really is. For example, if we look at an oar partially submerged in water we see the oar, but it might look bent. Illusions can naturally be sorted together with hallucinations, since both of them are misleading. On the other hand, they can also be naturally sorted with veridical perceptions, since in both cases the perceiver perceives an external object. Illusions will not figure in my argumentation for brain theory, neither do they figure in any objections against brain theory. Because of this, I will not consider what theories say about illusions when classifying theories of phenomenal qualities below.

We can make a coarse distinction between four types of general positions on phenomenal qualities in the good case. First, there are those who take them to be internal: properties instantiated in the brain or non-physical mental objects. Let us denote this position by “I.” Second, some take them to be external: properties of external perceived physical objects. Let us denote this position by “E.” Third, a few think phenomenal qualities are experienced, but not instantiated, in veridical perceptions. Let us denote this position by “N.” Fourth and finally, it is possible to hold that we do not even experience phenomenal qualities in veridical perceptions. Let us denote this last position by “X.”

In the bad case we have one option less. Here phenomenal qualities cannot be properties of external perceived physical objects since, by definition, the subject of the bad case is having a hallucination and so does not experience external objects which the experienced qualities could belong to. Otherwise, the same positions on locations of the phenomenal qualities as in the good case are possible, which we can let the same letters denote. So there are three distinct positions on the phenomenal qualities in the bad case: they can be internal “I,” they can be uninstantiated “N,” or it could be the case that we do not experience phenomenal qualities in hallucinations “X.”

---

located in the brain. Thus, it is useful to look at what alternatives the brain theory says about the location of phenomenal qualities experienced in hallucinations in particular.

<sup>60</sup> Similar terminology is used by Williamson (2000) and Byrne and Logue (2009b).

In theory, positions on the qualities in the good case can be freely combined with positions on the qualities in the bad case. However, not all combinations are reasonable. If phenomenal qualities are internal in the good case there is no good reason to think they are not internal in the bad case.<sup>61</sup> Also, if the phenomenal qualities are uninstantiated in the good case there is no reason to believe that they are instantiated in the bad case.<sup>62</sup> By the same reasoning, if the phenomenal qualities are not even experienced in the good case there is no reason to think that they are experienced, or instantiated, in the bad case.

If we let pairs of letters denote combinations, where the first letter stands for the position on phenomenal qualities in veridical perceptions and the second letter stands for the position on phenomenal qualities in hallucinations this leaves us with the following six combinations: "II," "EN," "EX," "EI," "NN," and "XX." In the following sections I will introduce and discuss these positions.

## 4.2 II theories

II theories say that both in hallucinations and veridical perceptions, phenomenal qualities are experienced and that these qualities are internal, in the sense that they are either located in the brain, or located in non-physical mental objects. Thus, on II theories, phenomenal qualities are never instantiated in external objects; as we might naïvely take them to be.

---

<sup>61</sup> This is for two reasons. First, there can be special reasons to want to resist internal qualities in the good case which do not apply to the bad case. For example, one might think that internal qualities in the good case prevent direct perception or goes against common sense. But there are no analogous reasons to think that internal qualities in the bad case are problematic. Also, the positive reasons for thinking that the experienced qualities are internal are if anything stronger in the bad case, not weaker. Thus, if we accept that phenomenal qualities are internal in the good case we should reasonably also accept that they are internal in the bad case.

Second, there are arguments with the conclusion that qualities in hallucinations are internal that would not directly apply to the good case. An example of this is the first part of the argument from hallucination, which concludes that qualities are internal in hallucinations. For a modern presentation of this, see Robinson (1994). In the light of such arguments it might be reasonable to hold that qualities are internal in the bad case but not the good, however the position that they are internal in the good case but not the bad would be unreasonable.

<sup>62</sup> This is because all reasons for thinking phenomenal qualities are instantiated that apply in the bad case also apply in the good case. For example, if one believes that when a subject experiences phenomenal qualities these qualities must also be instantiated phenomenal qualities would be instantiated in both veridical perceptions and hallucinations. However there might be additional reasons that apply in the good case, such as that we perceive external objects and that we ordinarily think these instantiated phenomenal qualities we experience in veridical perceptions. Thus, it can be a reasonable position to say that phenomenal qualities are instantiated in veridical perceptions but uninstantiated in hallucinations, whereas the reverse position that claims that phenomenal qualities are instantiated in hallucinations but uninstantiated in veridical perceptions is not a reasonable position.

Naturally, there are two broad categories of II theories: those that locate phenomenal qualities in the brain and those that locate phenomenal qualities in non-physical mental objects. Brain theory itself, which, as I mentioned in the introduction, roughly says that phenomenal qualities are physical properties instantiated in the brain is a clear example of a theory in the first category. As we shall see below, “sense data theory” is a clear example of a theory in the second category.

Next, I will look at four different theories of perception, which have associated theories of the phenomenal qualities that can plausibly be taken to exemplify versions of II theories: sense data theory, adverbialism, qualia theory and “internalist representationalism.” Let us look at these in turn.

Sense data theory is a classic theory of perception which has been around at least since the beginning of the last century. Although it seemed to be the dominant theory until the 1950s or 1960s, it is relatively unpopular today.<sup>63</sup> According to Fish (2010), sense data theory about perception analyzes visual experience as follows:

A subject S has a visual experience as of a property F if and only if: S senses an F sense datum, D (Fish 2010, 16).

Experiences in modalities other than vision have similar analyses. For examples of sense data theories, see Jackson (1977) and Robinson (1994).

The sense data themselves are today most often simply defined as non-physical objects which we are directly aware of in perception, regardless of whether it is veridical perception or hallucination.<sup>64</sup> Usually they are also taken to be logically private to their subject, which means that a certain sense datum token is only observable by the subject who experiences it.<sup>65</sup> As they are also supposed to be the bearers of the phenomenal qualities, sense data theory is both a theory of perception and a theory of phenomenal qualities.<sup>66</sup>

---

<sup>63</sup> According to the PhilPapers survey only about 3.1% of surveyed faculty members lean towards or accept sense data theory (Bourget and Chalmers 2009).

<sup>64</sup> When the term sense data was introduced at the beginning of the last century some, like Moore (1953, 30) did, used “sense data” as a neutral term for the objects of perception. In that use of the term sense data could be physical. For example this was the case for e.g., Russell (1914). On the other hand they were neither on the mind side nor on the matter side according to e.g., Price (1950). However, since then this usage has changed and sense data are now taken to be non-physical objects.

<sup>65</sup> Even if another person would experience an indistinguishable sense datum at the same location and time this would not be the same token sense datum but just a different token sense datum of the same type.

<sup>66</sup> See Huemer (2011).

It is thus clear what sense data theory says about phenomenal qualities in hallucinations: they are instantiated in non-physical objects, sense data.

*Sense data theory:* Phenomenal qualities experienced in hallucinations are instantiated in non-physical objects, sense data.

It says that phenomenal qualities are properties of non-physical objects: sense data. I have named this theory the *sense data* theory because of its close relation to the sense data theory of perception. We can now clearly see that sense data theory about phenomenal qualities is an II theory, since it does not distinguish between veridical perceptions and hallucinations.

The second theory of perception which can exemplify II theories of phenomenal qualities is adverbialism. It is a theory of perception which gives an adverbial analysis of visual experience, and similar adverbial analyses of experiences from the other perceptual modalities:

A subject S has a visual experience as of a property F, if and only if S senses F-ly (Fish 2010, 37).

The “F-ly” part of the analysis is the source of the theory’s name since it is the result of modifying the term ‘F’ into its adverbial form ‘F-ly’. For example, if one has an experience as of redness adverbialism says that one *senses redly*. A motivation for this might be that unlike sense data theory’s analysis of an experience as of a property F the adverbialism analysis avoids an overt commitment to *objects* that are F. It is instead only committed to subjects sensing in certain ways. Thus it might be able to do away with mental objects, such as sense data.<sup>67</sup>

It is not obvious what adverbialism, as a theory of perception, implies for phenomenal qualities. Some interpret adverbialism as ascribing qualities to experiences, as modifications of the experiences.<sup>68</sup> Adverbialism interpreted this way takes the phenomenal qualities to be internal both in the good case and in the bad case. Thus, one version of adverbialism belongs to the category II. Furthermore, this version of adverbialism is compatible with brain theory, assuming that experiences are brain states, since it would entail that phenomenal qualities are instantiated in the brain.

---

<sup>67</sup> Crane (2011) discusses this as a motivation for the arguments of Ducasse (1942) and Chisholm (1957) for adverbialism over sense data theory.

<sup>68</sup> See Fish (2010, 38) and Crane (2000, 177).

However some adverbialists, such as Cornman (1971), deny, at least for their own versions of adverbialism, that the adverbial terms ascribe qualitative properties to experiences and so give another interpretation of adverbialism.<sup>69</sup> Presumably these adverbialists instead either think that the phenomenal qualities, at least in the good case, belong to the external environment, or they think that there are no instantiated phenomenal qualities in either the good or the bad case. Since adverbialists give the same analysis of the good and the bad case they cannot say that the phenomenal qualities are internal in the bad case and external in the good case. This leaves EN/EX and NN/XX as the possibilities for which category of theories this version of adverbialism might belong to.

Let us move on to the third theory of perception that can take phenomenal qualities to be internal both in the good and the bad case, and so exemplify II. Sometimes a theory called “qualia theory” is mentioned among theories of perception.<sup>70</sup> I have found few clear statements of it as a theory of perception in the literature.<sup>71</sup> On one interpretation, qualia theory says the following. In both veridical and non-veridical perception there are instantiated qualia. They correspond to experienced sensible qualities in the sense that for each type of sensible quality there is a corresponding quale. For example, corresponding to redness there are redness qualia. We are also at least sometimes aware of these qualia, for example, when we reflect on our experience. Furthermore these qualia belong to experiences rather than to external objects. Finally, the qualia constitute phenomenal character and are intrinsic and non-representational properties.<sup>72</sup>

It is not completely obvious what qualia theory says about phenomenal qualities. On a first interpretation there is only one set of qualities, the phenomenal qualities, which are identical to the qualia. A consideration that speaks in favor of this interpretation is that phenomenologically we only seem to be aware of one set of properties, not two. Thus it would seem likely that qualia theorists do not distinguish qualia from phenomenal qualities. Furthermore, a more important consideration is that qualia theorists take qualia to constitute phenomenal character. Given the assumption discussed in 2.2 that the term ‘qualia’, when referring to the constituents of phenomenal character, typically refers to phenomenal qualities, we have a perhaps stronger reason to believe that at least some qualia theorists take qualia to be phenomenal qualities.

---

<sup>69</sup> See Cornman (1971, 271).

<sup>70</sup> For example, it is included among the alternatives views on perception in the PhilPapers survey, where 12.2% of surveyed faculty members accept or lean towards qualia theory (Bourget and Chalmers 2009).

<sup>71</sup> But Block (1996) might be an example.

<sup>72</sup> For these claims see Loar’s description of qualia theories in Loar (2003).

On this interpretation qualia theory claims that in both the good and the bad case phenomenal qualities belong to experiences and are thus internal, which implies that qualia theory belongs to the category II. More specifically, it would claim, just as brain theory, that the phenomenal qualities are instantiated in the brain.

On a second interpretation of what qualia theory says about phenomenal qualities, there are two sets of qualities. On the one hand, there are the phenomenal qualities, which belong to external objects, if anything. But on the other hand, there is also a second distinct set of qualities that belong to experiences, the qualia. What favors this second interpretation is that it seems relatively clear that we are generally directly aware of phenomenal qualities, whereas qualia theorists typically claim that we need a special procedure to become directly aware of qualia, or at least that we lack direct awareness of qualia.<sup>73</sup> On this interpretation qualia theory would belong to the category EN or the category NN. Just as for adverbialism above, since qualia theory gives the same analysis for the good and the bad case this type of qualia theory could not say that phenomenal qualities experienced in the bad case are internal and could thus not belong to EI. Thus, either the phenomenal qualities are external in the good case and not instantiated in the bad case or they are not instantiated in either case.

Finally, let us look at a fourth theory of perception, “internalist representationalism.” This theory is a version of the more general theory of perception “representationalism,” which I will discuss at length in section 4.5 below. “Internalist representationalism” says that the phenomenal qualities are representational properties of brain states. Representational properties are properties that represent some representational content. For example, a newspaper has a certain representational content and it represents this content in virtue of the representational properties of the letters written on it.

*(Internalist representationalism)* Phenomenal qualities are representational properties of brain states.

Internalist representationalism is described and discussed by Thompson (2008), who discusses what he calls “vehicle-based representationalism,” which is a view that seems to be a rather clear case of internalist representationalism. According to the view he discusses, “phenomenal properties,” which plausibly can be taken to be phenomenal qualities, are

---

<sup>73</sup> For example of such a procedure see Loar (2003) and for an example of a type of direct awareness which we cannot stand in to qualia, see Stoljar (2004) sect. 6.3.

identical to representational properties of brain states. This is the case both in veridical perception and hallucination:

According to vehicle-based representationalism, a phenomenal property is the property of representing *p* rather than the represented property *p* itself. ... phenomenal character supervenes on intentional content ... [it] is a property of the vehicle of content – the property of having a particular content (Thompson, 2008, 395).

Vehicle-based representationalism identifies phenomenal character with the property of *representing* that *p*, rather than with the intentional content *p*. Prima facie, the property of representing that *p* is a property that is really instantiated during hallucination... On this view, phenomenal properties are properties of experiences rather than properties of external objects (Thompson, 2008, 401).

The theories of Block (1996) and Loar (2003) might be examples of internalist representationalism. This position is clearly an II theory, belonging to the broad category of II theories that claim that phenomenal qualities are instantiated in the brain.<sup>74</sup>

### 4.3 EI theories

We can now move on to the second category of theories, EI. It contains theories which claim that we experience phenomenal qualities in both the good and the bad case, and that these experienced phenomenal qualities are external in the good case and internal in the bad case.

A motivation for why one would want to hold an EI view is that if qualities are external in the good case we can easily do justice to our naïve view of perception, and if qualities are internal in the bad case we can easily explain how we can experience phenomenal qualities in hallucinations.

On the other hand, generalizing arguments, which claim that phenomenal qualities have the same location in veridical perceptions as in veridical perceptions, pose a problem for EI views. Furthermore, if the qualities in the

---

<sup>74</sup> Note that internalist representationalism is also compatible with brain theory. In fact, as I have defined it, internalist representationalism is a more specific version of brain theory. Where brain theory says that phenomenal qualities are properties of the brain internalist representationalism says that the phenomenal qualities are representational properties of the brain.

bad case are located in non-physical objects, the implication that physicalism is false might also be taken to be problematic.

An EI theory can provide any internalist account of phenomenal qualities experienced in hallucinations that were discussed above, such as brain theory, sense data theory, adverbialism, qualia theory or internalist representationalism. Thus, it can agree with brain theory about the location of phenomenal qualities in hallucinations. The difference between EI and II theories is that the former only uses the internalist account for hallucinated phenomenal qualities, and not for phenomenal qualities experienced in veridical perceptions. For example, an EI theory may claim that phenomenal qualities experienced in veridical perceptions belong to external objects, whereas phenomenal qualities experienced in hallucinations belong to sense data.<sup>75</sup>

One of these EI theories will be argued against separately in my argumentation for brain theory below: EI brain theory. It says that the phenomenal qualities experienced in hallucinations are instantiated in the brain, but that phenomenal qualities in veridical perceptions have a different location. We can define it as follows:

*EI Brain theory:* All phenomenal qualities experienced in hallucinations are instantiated in the brain. Phenomenal qualities experienced in veridical perceptions are not instantiated in the brain.

This view does not provide an alternative to brain theory for the phenomenal qualities experienced in hallucinations, but it still is a distinct alternative to brain theory, since it disagrees with brain theory on the location of phenomenal qualities experienced in veridical perceptions. If we exclude brain theory versions of the unreasonable views XI and NI, it also follows that EI brain theory subsumes all alternative positions that do not disagree with brain theory about hallucinations.<sup>76</sup>

---

<sup>75</sup> Byrne and Logue (2009b) discuss a version of disjunctivism about perception which claims that in veridical perception the perceiver stands in a perceptual relation to an external object whereas in hallucination she stands in some other relation, such as “acquaintance” in Russell’s (1982) sense, to a non-physical sense data. That would be a version of EI-disjunctivism, assuming that the phenomenal qualities experienced in the hallucination also belong to the sense data the perceiver is acquainted to. Alston’s (1999) theory of perception might be another example of EI-disjunctivism.

<sup>76</sup> Another theoretically possible view is an II view which claims that phenomenal qualities experienced in hallucinations are properties of the brain but phenomenal qualities experienced in veridical perceptions are properties of sense data. However, as no one holds such a view and there are no good reasons to hold it I will not consider it. Furthermore, if brain theory is correct about hallucinations then II-disjunctivism is also ruled out by my generalizing arguments in Chapter 9.

Let us define *phenomenal quality disjunctivism* as any view which gives fundamentally different accounts of phenomenal qualities experienced in the good case and phenomenal qualities experienced in the bad case.<sup>77</sup> Thus, all EI theories are versions of phenomenal quality disjunctivism, since they give fundamentally different accounts of the qualities in the good and the bad case. Below, when looking at EX and EN theories, we will see that such theories can also be versions of phenomenal quality disjunctivism, although there are also EN theories which are not versions of it.

Let us proceed to looking at a theory of perception which can exemplify EI: disjunctivism. This is a relatively new and hotly debated theory, which claims that we should give fundamentally different accounts of hallucinations and veridical perceptions.<sup>78</sup> It is named “disjunctivism” because it analyzes experience as a disjunction: an experience is *either* a veridical perception *or* a hallucination. For example, according to Fish (2010, 88) the disjunctivist gives the following account of visual experiences:

A subject S has a visual experience as of a property F if and only if:  
\*either S perceives an F,  
\*or S has a hallucination of an F.

A disjunctive analysis rejects the claim that indistinguishable hallucinations and veridical perceptions share a fundamental common kind.<sup>79</sup> This is a way of saying that even if a non-veridical perception is indistinguishable from a veridical perception, the two events are not fundamentally the same kind of thing: they have different essences.<sup>80</sup>

---

<sup>77</sup> I choose the name “disjunctivism” because of the similarity to the disjunctive theory of perception, which I consider below.

<sup>78</sup> According to the PhilPapers survey, 10.9% of surveyed faculty members leans towards or accepts disjunctivism (Bourget and Chalmers 2009). Disjunctivism has gained a lot of attention recently. For example, two volumes on disjunctivism, Byrne and Logue 2009a and Haddock and Macpherson 2008, were published recently, one collecting classical readings on disjunctivism and the other collecting contemporary readings on disjunctivism.

<sup>79</sup> On the notion of a fundamental kind in this context, see Martin (2004).

<sup>80</sup> Note that this general disjunctive analysis does not in itself say anything about the nature of veridical perceptions or the nature of hallucinations: it just says that they are different. Thus, it can be combined with any account of the disjuncts. In practice, it is exclusively combined with a “naïve realist” account of veridical perception. That theory is called “naïve” because it claims that things in perception are as they supposedly seem to naïve subjects. One way to give a precise definition of naïve realism, which is used by disjunctivists, is to say that perceived external objects are constituents of veridical perceptions. Now we see that if naïve realism is true, there must be a fundamental difference between indistinguishable hallucinations and veridical perceptions, since perceived external objects cannot be constituents of hallucinations. This can lead the naïve realist to reject the claim that hallucinations and veridical perceptions share a fundamental common kind, and thus turn to disjunctivism. Note that this way of giving a precise definition of naïve realism is not entirely uncontroversial, since some theories of perception which are not disjunctivist theories might also want to claim to be versions of naïve realism.

There are many different types of disjunctivism.<sup>81</sup> The one that is closest related to the view I defined as disjunctivism about phenomenal qualities is *disjunctivism about phenomenal character*. The latter view says that indistinguishable veridical perceptions and hallucinations differ in their phenomenal character, and thus also their qualia. For example, such a disjunctivist might claim that qualia in veridical perceptions are properties of external objects, whereas indistinguishable hallucinations either lack qualia, or have qualia that are properties of something else.

Given the assumption discussed in 2.2, that the term ‘qualia’ typically refers to phenomenal qualities, it is likely that many of the disjunctivists about phenomenal character take the qualia which constitute the phenomenal character to be phenomenal qualities. Thus, it seems likely that disjunctivists about phenomenal character also are disjunctivists about phenomenal qualities.

#### 4.4 EX theories

Let us proceed to the third category of theories about phenomenal qualities, EX. This contains theories which claim that phenomenal qualities are external in the good case, but deny that any phenomenal qualities are experienced in the bad case. Since these theories deny that we experience phenomenal qualities in one of the cases, but not in the other, they give fundamentally different accounts of phenomenal qualities in the good case and the bad case. Thus, these theories are all versions of phenomenal quality disjunctivism, as defined above.

The theory types we considered above, II and EI, both agreed on the experiencing question, instead they differed only on what location they gave phenomenal qualities experienced in hallucinations. EX however, stands out because of its negative answer to the experiencing question: the denial of the claim that we experience phenomenal qualities in hallucinations.

Now, what does it mean to *experience* phenomenal qualities more precisely? Is it possible to explain this expression further? Unfortunately, it seems to me that the experiencing-relation in which we stand to phenomenal qualities when we experience them is primitive, and thus not analyzable. However,

---

<sup>81</sup> For overviews of the different types of disjunctivism, and a general introduction to disjunctivism, see Byrne and Logue (2009b) and Soteriou (2010).

one thing that can be said to illuminate the relation is that the claim that we experience phenomenal qualities in an experience seems to be equivalent to claim that the experience *has phenomenal character*. We may also be able to improve our understanding of the relation by looking at specific theories which deny that we experience phenomenal qualities in hallucinations, such as EX theories. For example, the EX claim that we do not experience phenomenal qualities in hallucinations seems to imply that hallucinations are not conscious experiences. Finally, note that the statement that phenomenal qualities are experienced in hallucinations does not presuppose that the qualities are instantiated in a perceived object. Thus, that statement does not exclude positions of the types EN and NN, which claim that phenomenal qualities in hallucinations are uninstantiated.

If hallucinations are not conscious experiences, what are they then? One way to develop an EX view is to claim that hallucinations, rather than being conscious experiences, are purely cognitive states, much like beliefs. When we have a hallucination, we believe that we experience phenomenal qualities; however, these beliefs are false. In virtue of having such beliefs, however, we might at least be said to be related to phenomenal qualities in hallucinations, although this is a relation of cognizing the qualities, rather than an experiencing relation. As this developed view can seem more plausible than the bare EX claim that we do not experience phenomenal qualities in hallucinations, I will focus on this version of EX below in my argumentation, when arguing against EX theories.

Just as for the EI theories above, the theories of perception which might exemplify EX belong to the disjunctivism family. One possible example of an EX theory is the disjunctivist theory of perception of Fish (2008, 2009):<sup>82</sup>

[P]ure hallucinations that take place in the absence of any background experience of the world... will therefore not have an acquaintance-based phenomenal character ... [W]e should also reject the claim that pure hallucinations acquire phenomenal character of some different kind from an alternative source. This is why the theory I present here is, as has already been foreshadowed, *eliminativist* about hallucinatory phenomenal character (Fish 2010, 93).<sup>83</sup>

Fish accepts that hallucinations seem to have a phenomenal character, but according to him this is just because they have the same cognitive effects as

---

<sup>82</sup> ). Sturgeon (2008, 142) suggests that in many “dialectical settings” a disjunctivism which denies that hallucinations have phenomenal character is the best version of disjunctivism and perhaps even the best theory of visual experience *simpliciter*.

<sup>83</sup> As we will see in Chapter 5, when I use the term ‘hallucination’ I talk about pure hallucinations.

indistinguishable veridical perceptions would have. Although a hallucination does not have phenomenal character it makes us claim, think and believe the same things as an indistinguishable veridical perception would make us claim, think and believe. Thus, although the hallucination does not have phenomenal character, it causes us to believe that it has phenomenal character:<sup>84</sup>

[I] develop a theory of hallucination that claims that hallucinations are experiences that lack phenomenal character, but that nonetheless lead subjects to believe that they see something Fish (2009, 80).

Because [a hallucination] produces the same cognitive effects as a veridical perception, a suitable sophisticated subject would still *believe* that it has phenomenal character, *think* that there is something it is like for him to hallucinate in such a way, and *claim* that he is having an experience of a certain kind, despite being mistaken. To paraphrase Armstrong the phenomenal character of a hallucination is simply a ghost generated by my belief that I am seeing something Fish (2009, 98).

Although Fish only explicitly discusses phenomenal character I take it that his position also implies that we do not experience any phenomenal qualities in hallucinations. The reason is that according to Fish hallucinations *lack* phenomenal character. It is not only that the phenomenal character of hallucinations is constituted by non-existing, or uninstantiated qualities, as some theories we shall encounter in the next section claims, but Fish states very clearly that hallucinations lack phenomenal character. Thus, we have reason to believe that his theory is an EX theory.

Moving on, Martin (2002, 2004, 2006) might provide another example of an EX theory. According to his special brand of disjunctivism about perception, so called “negative disjunctivism,” nothing more can be said about the mental characteristics of a hallucination than that it is indiscriminable from a veridical perception:

When it comes to a mental characterization of the hallucinatory experience, nothing more can be said than the relational and epistemological claim that it is indiscriminable from the perception (Martin 2004, 72).

According to Martin the only facts there are about hallucinations are that they stand in epistemic relations of indiscriminability to veridical perceptions. If phenomenal qualities were experienced in hallucinations something more could be said about hallucinations than the relational claim above, it seems. For example, we could say that in hallucinations

---

<sup>84</sup> See Fish (2009, 97-99).

phenomenal qualities are experienced. So on one interpretation of his negative disjunctivism it denies that phenomenal qualities are experienced in hallucinations, making it an EX theory.

On the other hand, in later papers he explicitly denies that hallucinations lack phenomenal character. Rather than being a part of this position, Martin claims, that is an unappealing consequence that his opponents try to show follows from his negative disjunctivism, but which does not follow:

(III) For certain experiences as of a white picket fence, namely, causally matching hallucinations, there is no more to the phenomenal character of such experiences than that of being indiscriminable from corresponding visual perceptions of a white picket fence as what it is. (Martin 2006, 369)

The debate here is not over whether both sides should agree that the subject genuinely has sense experience in the case of causally matching hallucination in contrast to the case of post-hypnotic suggestion. Both sides should agree with that. The question is whether the disjunctivist can show that one can coherently claim this while also affirming (III) (Ibid. 377).

Going by this it seems his position says that *what it is* for a hallucination to have phenomenal character is that it stands in a certain epistemic relation to a veridical perception. Thus, assuming that when Martin talks about “phenomenal character” he talks about that which is constituted by phenomenal qualities, his theory is not an EX theory. Rather, on this interpretation his disjunctivism is a version of the EN family, which is discussed in the next section.

#### **4.5 EN theories**

Let us move on to the fourth category, EN. It contains theories which claim that we experience phenomenal qualities in both veridical perceptions and hallucinations, and take these experienced phenomenal qualities to be external in the good case, but uninstantiated in the bad case.

The negative claim that phenomenal qualities are not instantiated does not give them a very precise location. Because of this, I will distinguish two broad categories of EN theories, that differ in the precise location they give

phenomenal qualities experienced in hallucinations, and then distinguish two more subcategories within the second of these.

First, let us define ‘Property-ism’ as follows:

*Property-ism:* Phenomenal qualities are experienced in hallucinations and these qualities are uninstantiated properties themselves.

Here we have our third alternative location for the phenomenal qualities experienced in hallucinations, as uninstantiated properties. The term “uninstantiated” should in this context be understood referring to properties which are not instantiated in an object perceived by the subject of the experience, nor in any object internal to the subject. Thus, a whether phenomenal quality is uninstantiated in this sense is relative to a subject and an experience:

A phenomenal quality Q experienced in an experience E had by a subject S is *uninstantiated* relative to E if it is not instantiated in any object perceived by S or in any object internal to S.

What uninstantiated properties does the Property-ist think that the phenomenal qualities are identical to? The most reasonable choice would be the *external candidates*, the external properties that are the best candidates for identification with the phenomenal properties. For example, a phenomenal redness quality might be identical to the external surface property “external redness.” A phenomenal squareness quality might be identical to a geometrical property, “external squareness.”

An example of a philosopher who might plausibly be interpreted as a property-ist is Michael Tye:<sup>85</sup>

[E]ven if [a] subject is hallucinating, redness itself exists. The subject is aware of the quality, redness, and undergoes an experience that (mis)represents that something in the vicinity of the viewer has it. This is what happens in the case of experiencing a red after-image, for example. There is no filmy, red spot floating in space, but to the subject it seems that there is. Here the experienced quality, red, has no bearer (Tye 2005, 169)

Tye here claims that the subjects of hallucinations are aware of phenomenal qualities but that these qualities are uninstantiated. Furthermore, he makes

---

<sup>85</sup> Other plausible examples of property-ists include Dretske (1993), Forrest (2005) and Johnston (2004).

the more specific claim that the qualities the subject of the hallucination experiences is the properties themselves, which exist. Thus, on this view, the phenomenal qualities we experience in hallucinations do in fact exist, since they are the uninstantiated properties themselves.

Second, we might define a distinct EN position, which unlike Property-ism denies that the phenomenal qualities experienced in hallucinations exist, as follows:

*Non-existence-ism:* We experience phenomenal qualities in hallucinations but these qualities do not exist.

A clear example of a quality non-existent-ist is William Lycan. When describing “the representationalist theory of perception”, a theory which I will explain in further detail below, and which is a theory that he himself also defends, Lycan claims that according to representationalism the qualities we experience in hallucinations are non-actual material things:

Suppose Bertie is experiencing a green after-image as a result of seeing a red flash bulb go off; the greenness of the after-image is the quale. ...To sharpen the problem, suppose there is no green physical object in Bertie's visible environment either. ... What about Bertie's green after-image? On the representationalist (sometimes "intentionalist") analysis, for Bertie to experience the green after-image is for Bertie to be visually representing a green blob located at such-and-such a spot in the room. Since in reality there is no green blob in the room with Bertie, his visual experience is unveridical; after-images are illusions. *The quale, the greenness of the blob, is (like the blob itself) a nonactual intentional object* (my italics). Of course, in cases of veridical perception, the color and the colored object are not merely intentional contents, because they actually exist, but they are still intentional objects, representata. (In defending his sense-data, Russell mistook a nonactual material thing for an actual immaterial thing) Lycan (2008).

Here Lycan describes representationalists as claiming that phenomenal qualities experienced in hallucinations are non-existing objects. Given our definition above this implies that representationalism is a version of Non-existence-ism, according to Lycan.

When describing his own theory he is even more specific about the nature of phenomenal qualities experienced in hallucinations. They are possibilities, merely possible things that are located in other possible worlds:

Being found problematic when construed as genuinely inhering in ordinary physical objects, the secondary qualities were kicked upstairs into the mind, and made into properties of sense-data or at any rate modes of sensing. I now propose to kick them out of the mind, but not precisely back downstairs - rather, into neighboring possible worlds (Lycan 1987, 90-91).

Another possible example of a non-existent-ist is Gilbert Harman<sup>86</sup>.

However, few representationalists are as explicit as Lycan's claim that the phenomenal qualities in hallucinations do not exist. A more common statement is that they are properties of non-existing objects. This could either be taken to mean the same thing or it could mean that these phenomenal qualities are uninstantiated properties, as the Property-ist claims.

Furthermore, it seems to me like the Non-existence-ism itself is ambiguous and can be interpreted in at least two ways. First, we might read the claim on its face, as saying that the experienced qualities just do not exist, *simpliciter*. Second, we might think it says that the experienced qualities have a peculiar mode of being, that they do not count as 'existing' but still "have being" in some sense.<sup>87</sup> This gives rise to two distinct sub-positions of Non-existence-ism:

*Nothing-ism:* Phenomenal qualities in hallucinations are experienced but do not exist, *simpliciter*.

*Meinongianism:* Phenomenal qualities in hallucinations are non-existing entities which "have being."

These two subpositions provide our fourth and fifth alternative locations for the phenomenal qualities experienced in hallucinations: as not existing *simpliciter* and as not existing but "having being."

Finally, I will look at what theories of perception can exemplify EN positions on the phenomenal qualities. In doing this I will first look at the

---

<sup>86</sup> He claims that the qualities we experience in hallucinations are qualities of non-existent things. That claim could be consistent either with Property-ism or Non-existence-ism. On the one hand it could express a version of Property-ism, since uninstantiated qualities in one sense could be said to be qualities of non-existing objects. On the other hand it could express Non-existence-ism since non-existing qualities belong to non-existing objects. Perhaps the latter interpretation is most plausible in this case since Harman here only talks about the qualities as belonging to non-existing objects, without mentioning them as uninstantiated. Thus, Harman might be another example of a Non-existence-ist philosopher.

<sup>87</sup> On the idea of objects that *are* but do not exist, see Reicher (2012).

representationalist theories of perception, which have already been mentioned, and then, finally, briefly look at disjunctivist theories.

Representationalism is currently the most popular theory of perception.<sup>88</sup> A core idea among representationalists is that perception is a form of representation. There are many different versions of representationalism, or intentionalism as it is sometimes called. Let us use the following taxonomy to distinguish them. *Weak representationalism* is the claim that every perception necessarily has representational content; representational content is a condition of satisfaction on experiences that for each experience is either fulfilled or not fulfilled.<sup>89</sup> Thus, just as beliefs or sentences can be true or false, experiences can be either correct or incorrect.<sup>90</sup> For example, an experience as of a red square might have the representational content that there is a red square at a certain location in front of the perceiver. The content is satisfied and the experience correct if there is such a red square in front of the perceiver. *Weak representationalism* is a sort of minimal representationalist claim and so is included in all forms of representationalism.<sup>91</sup> The position is false if experiences lack such a condition of correctness, for example if they only *present* perceptual items without representing anything.

*Strong representationalism* states that in addition to the former claim, there is a change in phenomenal character of experiences only if there is a corresponding change in the contents of the experiences.<sup>92</sup> This means that

---

<sup>88</sup> According to the PhilPapers survey about 31.4% of surveyed faculty members lean towards or accept representationalism (Bourget and Chalmers 2009).

<sup>89</sup> On this definition of representational content see Chalmers (2004) and Chalmers (2010).

<sup>90</sup> Denying this claim means that experiences do not have satisfaction conditions. For example Travis (2004) claims that experiences just present the perceived environment, rather than representing anything as being "thus and so." Experiences might sometimes be misleading, in the sense that we draw the wrong conclusions from them but from this it does not follow that they have satisfaction conditions. Presumably it is still possible to distinguish veridical perceptions from illusions and hallucinations, just not by their satisfaction conditions. See also Siegel (2011).

<sup>91</sup> See Fish (2010, 66) and Crane (1998, 233) for discussions of weak representationalism. Crane is an example of a weak representationalist.

<sup>92</sup> As Fish (2010) uses the term "strong representationalism" there is a change in the phenomenal character of experiences *if and only if* there is a corresponding change in the contents of the experiences. However, this condition seems too restrictive, since some versions of representationalism that postulate an intuitively far stronger connection between phenomenal character and representational content than weak representationalism do not endorse the "if-direction" of the "if and only if claim." For example according to Chalmers' (2004) representationalism there is a certain type of content, Edenic content, for which there is a change in phenomenal character if and only if there is a change in that content. But at the same time Chalmers claims that experiences also have additional content, which can change without changes in phenomenal character. For example, experiences can have Russellian, object-involving content. These are contents that represent certain objects as having certain properties. Such Russellian contents are satisfied only if a certain object O has a certain property P. According to Chalmers, an experience with the same phenomenal character can be directed at a different object, O', and thus be satisfied only if the distinct object O' has P. This provides an example of two experiences with the same phenomenal character, but with different Russellian contents, and so of a case where we can have a difference in content without a difference in phenomenal character.

which phenomenal qualities we experience in an experience cannot vary independently of the experience's representational content. Thus the phenomenal qualities we experience in an experience depend on the representational content of that experience in the sense that the former does not change if the latter do not change. Some versions of representationalism claim that the ground of this dependence-relation is that the representational content of an experience explains its phenomenal character.<sup>93</sup> A particular sort of strong representationalism is reductive externalist representationalism, or as I will call it, *externalist representationalism*.<sup>94</sup> It is reductive in the sense that it identifies the qualia in an experience with properties included in the content of that experience. It is externalist in the sense that it claims that the properties included in the content of experiences generally are properties of external objects:

On wide representationalism, qualia (like meanings) ain't in the head. The classic, Cartesian-based picture of experience and its relation to the world is thus turned upside down. Qualia are not intrinsic qualities of inner ideas of which their subjects are directly aware, qualities that are necessarily shared by internal duplicates however different their environments may be (Tye 2009).

For example, say we have an experience of a red square. According to externalist representationalism, the experienced phenomenal redness and phenomenal squareness in this experience are properties of external objects which are represented by the experience; in this case perhaps a surface reflectance property and a geometrical property.

Both weak and strong representationalism are compatible with brain theory. The former compatibility follows since brain theory brain theory is compatible with experiences having representational content, as we saw above in section 3.3. The latter compatibility follows since brain theory is not committed to any particular claim about the relation between phenomenal qualities and representational content. For example, assume the following three things, (1) that phenomenal qualities are always instantiated in the brain, (2) that the representational contents of experiences are completely determined by what phenomenal qualities are experienced in them, and (3) that only phenomenal qualities constitute phenomenal character. In this case there is a change in phenomenal character only if there is a change in phenomenal character, so that both brain theory and strong

---

<sup>93</sup> Some other representationalists think that the representational content of an experience depends on the phenomenal character of said experience and hold that experiences have representational content in virtue of their phenomenal character. Their theories are typically qualia theories, in addition to being representationalist theories. So there can be some overlap between representationalism and qualia theory. See Fish (2010, 68).

<sup>94</sup> Tye (2000) and Dretske (1993) are examples of externalist representationalists.

representationalism are true. Thus, the two theories are compatible, at least given the assumptions (1)-(3).

However, brain theory is incompatible with reductive externalist representationalism, assuming the latter's qualia are the phenomenal qualities. This is because according to such representationalism the phenomenal qualities are external. Contrary to brain theory, this means that if the qualities are instantiated anywhere they are not instantiated in the brain.

What about phenomenal qualities in hallucinations? Representationalists often say more about objects experienced in hallucinations than about properties experienced in hallucinations.<sup>95</sup> However, when arguing against sense data theory, a reductive externalist representationalist will typically emphasize that the objects of hallucinations do not exist. One way to justify this is to note that on representationalism perception is a form of representation. But the objects of representations need not exist. For example, if Ponce the Lyon is searching for the fountain of youth it does not following from this fact that there is an existing object that he is searching for. A classical statement of this line of thought is given by Harman:

Defenders of the sense datum theory argue for it by appealing to the so-called argument from illusion. This argument begins with the uncontroversial premise that the way things are presented in perception is not always the way they are. Eloise sees some brown and green. But there is nothing brown and green before her; it is all an illusion or hallucination. From this the argument fallaciously infers that the brown and green Eloise sees is not external to her and so must be internal or mental. ...

In order to see that such arguments are fallacious, consider the corresponding argument applied to searches; "Ponce de Leon was searching for the Fountain of Youth. But there is no such thing. So he must have been searching for something mental. This is just a mistake. From the fact that there was no Fountain of Youth, it does not follow that Ponce de Leon was searching for something mental. In particular he was not looking for an idea of the Fountain of Youth. He already had the idea. What he wanted was a real Fountain of Youth, not just the idea of such a thing (Harman 1997, 665).

The conclusion of the line of thought is that the objects of hallucinations do not exist: "[I]f Melvin hallucinates a pink elephant, the elephant that Melvin sees does not exist (Ibid. 668)."

---

<sup>95</sup> This is noted by Thompson (2008).

However, it is not obvious how experienced phenomenal qualities experienced enter into this account of hallucinations. Of those who mention them some, like Tye above, say that these are the uninstantiated properties themselves. Others, like Lycan above, say that the experienced qualities are properties of objects that do not exist or that the experienced qualities themselves do not exist. Below I will describe these possibilities in turn.

Finally, as we saw in the last section, some disjunctivist theories of perception might also belong to the EN category. These would be theories which agree that we experience phenomenal qualities in the bad case, unlike EX theories, and instead claimed that hallucinated qualities are experienced but uninstantiated. Such EN disjunctivists could use any of the more specific EN accounts of phenomenal qualities described above, such as Property-ism, Nothingism or Meinongianism. An EN disjunctivist can in fact give precisely the same account of the perceptual relation in hallucinations as a representationalist, as long as they do not also give a representationalist account of the perceptual relation in veridical perception. Furthermore, both of these theories also give the same answers to the location question, both in the good case and the bad case.

The main difference between EN disjunctivist theories and EN representationalist theories is, for our purposes, that the former is a version of quality disjunctivism whereas the latter is not. That EN representationalist theories are not quality disjunctivist follows because they give the same account of phenomenal qualities in both the good and the bad case, saying that the phenomenal qualities are represented properties. The reason these theories imply that phenomenal qualities have different locations in veridical perceptions and hallucinations is that the represented properties is instantiated in a perceived object in one of the cases, but not in the other. So despite giving the same account in both cases, the qualities are given different locations. EN versions of disjunctivism on the other hand, give fundamentally different accounts of the qualities in the good case and the bad case. For example, a version might claim that in the good case presented to the subject in a way that is more direct than mere representation, whereas in the bad case the experienced properties are merely uninstantiated represented properties.

## 4.6 NN and XX theories

Finally, I will briefly discuss the categories NN and XX. The former contains theories which claim that phenomenal qualities are neither *instantiated* in the good case nor in the bad case. The latter contains theories which go even further and say that we neither *experience* phenomenal qualities in the good case nor in the bad case.

One version of a NN theory is eliminative representationalism. An example of this position might be the representationalist theory of Pautz (ms), at least in the case of phenomenal color. Seemingly, according to his theory phenomenal qualities are colors. Pautz agrees that we experience phenomenal colors in both hallucinations and veridical perceptions. However, he claims, no objects are ever colored. On Pautz view, all color experience can be said to be illusory, in one sense. Thus, phenomenal colors experienced in perceptions are never instantiated, regardless of whether the perceiver is having an ordinary veridical perception or a hallucination. So his theory belongs to NN.

Another example of a NN theory could be the theory of perception of Chalmers (2010), at least under one interpretation of it. According to Chalmers' theory of the contents of perception, our experiences have so called "Edenic" properties as "phenomenal contents." By "phenomenal contents" he means representational contents of experiences that are determined by the phenomenal character of said experiences.<sup>96</sup> These Edenic contents are primitive, non-physical properties, and according to Chalmers they are never instantiated in the actual world. Thus, if we interpreted the theory so that phenomenal qualities are identified with such Edenic contents, it follows that we never experience instantiated phenomenal qualities. Since Chalmers does not deny that we experience phenomenal qualities in hallucinations, or veridical perceptions for that matter, this would make his theory an NN-theory.

Views of type XX deny that we ever experience phenomenal qualities in perceptions, regardless of whether we are having veridical perceptions or hallucinations. We might thus call positions of type XX "general eliminativism." These views seem rare. It is relatively uncontroversial that we experience phenomenal qualities in typical *veridical* perceptions. There are no obvious cases of philosophers who deny this. However, this lack of

---

<sup>96</sup> On Chalmers' view, experiences also have other non-Edenic properties as contents. However, only the Edenic properties are "phenomenal contents", in other words, representational contents perceptions have that are determined by the experiences' phenomenal character. On the definition of phenomenal contents, see Chalmers 2006.

obvious cases need not imply that no philosophers deny that we experience phenomenal qualities in typical veridical perceptions. The reason is that, as was noted in 2.2, philosophers sometimes write about phenomenal qualities in terms that can be interpreted to be about other things. Thus, the lack of obvious cases of philosophers who deny that we experience phenomenal qualities in typical veridical perceptions might be explained by such philosophers not speaking clearly about the phenomenal qualities rather than there not being any such philosophers.

Some things said by a few philosophers seem to imply that these philosophers assume the radical position that we do not even experience phenomenal qualities in veridical perceptions. A notable example is Dennett (1991):<sup>97</sup>

"There is no such thing as really seeming - over and above the phenomenon of judging in one way or another that something is the case... But what about the *actual* phenomenology? There is no such thing... *There seems to be phenomenology...* But it does not follow from this undeniable, universally attested fact that there *really is* phenomenology. (Dennett 1991, 364-366)

It seems like Dennett here claims that we do not experience phenomenal qualities. He denies that there is such a thing as phenomenology, and he denies that there is such a thing as "really seeming." Since he claims that it is undeniable and universally attested that there *seems* to be phenomenology, it seems likelier that he is referring to the phenomenal qualities when he uses the term "phenomenology" than that he is talking about some other further set of qualities that are internal. Thus, one natural interpretation of his claim is that in perception, veridical or not, there seems to be phenomenal qualities, but in reality there are no phenomenal qualities.

Still, other interpretations are also possible. Perhaps he believes that we experience phenomenal qualities but that these qualities are somehow unreal or non-existing, or perhaps he thinks that there are no phenomenal qualities in the brain, or in the mind, but that external objects might instantiate phenomenal qualities.<sup>98</sup>

Another possibility is that Dennett means something similar to what he seems to say in his "Quining Qualia" paper. There he speaks about "conscious experience" and "qualia", rather than "phenomenology", but it seems reasonable to assume that he is talking about the same things in both

---

<sup>97</sup> See also Rey (1983, 1988) and Wilkes (1988, 1995) for other possible examples of philosophers who deny that we experience phenomenal qualities in veridical perceptions.

<sup>98</sup> These are positions I will argue against in Chapter 7.

cases. In the paper he claims that while conscious experience exists, its nature is so unlike traditional conceptions of conscious experience, to such a high degree that it might be justified to make statements like “there are no qualia”:<sup>99</sup>

Which idea of qualia am I trying to extirpate? Everything real has properties, and since I don't deny the reality of conscious experience, I grant that conscious experience has properties. I grant moreover that each person's states of consciousness have properties in virtue of which those states have the experiential content that they do. That is to say, whenever someone experiences something as being one way rather than another, this is true in virtue of some property of something happening in them at the time, but these properties are so unlike the properties traditionally imputed to consciousness that it would be grossly misleading to call any of them the long-sought qualia. Qualia are supposed to be special properties, in some hard-to-define way. My claim--which can only come into focus as we proceed--is that conscious experience has no properties that are special in any of the ways qualia have been supposed to be special.

... My claim, then, is not just that the various technical or theoretical concepts of qualia are vague or equivocal, but that the source concept, the "pretheoretical" notion of which the former are presumed to be refinements, is so thoroughly confused that even if we undertook to salvage some "lowest common denominator" from the theoreticians' proposals, any acceptable version would have to be so radically unlike the ill-formed notions that are commonly appealed to that it would be tactically obtuse--not to say Pickwickian--to cling to the term. Far better, tactically, to declare that there simply are no qualia at all  
Dennett (1999, 620).

Here Dennett claims that qualia, in addition to constituting phenomenal character or conscious experience, are supposed to have certain other properties.<sup>100</sup> However, as he argues that nothing has the features he ascribes to qualia, it might be appropriate to say that there are no qualia, if

---

<sup>99</sup> Other philosophers, such as Chalmers (1996, 11), take “conscious experience” to refer to the phenomenal aspects of the mind. The term “phenomenology,” used in the context of denying that there is phenomenology, also refers to the phenomenal aspects of the mind. Thus it does not seem unreasonable to assume that Dennett means something similar in the case of both the expressions.

<sup>100</sup> Namely, the following properties:

So when we look one last time at our original characterization of qualia, as ineffable, intrinsic, private, directly apprehensible properties of experience, we find that there is nothing to fill the bill. In their place are relatively or practically ineffable public properties can refer to indirectly via reference to our private property detectors – private only in the sense of idiosyncratic (Dennett 1999, 639).

his arguments are correct, without denying the existence of phenomenal character.

Similarly, when he says that there is no phenomenology, and no such thing as really seeming, he might mean that there is conscious experience but that its nature is so unlike traditional conceptions of conscious experience that it is justified to say that there is no phenomenology or that there are no “seemings.” Thus, on this interpretation, he need not deny that there are phenomenal qualities in veridical perceptions.

## **5. Preliminaries to the argumentation**

### **5.0 Introduction**

In the first section of this chapter, I describe how and when my main argumentation gives reasons against the respective alternatives to brain theory. In the second section, I first define the core terms “hallucination,” “hallucination\*” and “veridical perception.” Then, I present the intermediary theses that we will use in our argumentation for brain theory. Finally, I present a brief argument that one of these theses is true.

### **5.1 How I argue against the alternatives to brain theory**

Prior to going through the parts of my argumentation, I will mention two restrictions. The first restriction is that I will not argue against XX-positions, in other words, positions which claim that phenomenal qualities are never experienced, not even in veridical perceptions. Instead, I will make the following assumption, which is incompatible with the truth of any XX theory:

*Realism assumption:* When a subject S has a veridical perception, S experiences phenomenal qualities.

The reason I assume this, rather than argue for it, is that arguing for it would require an extensive argumentation that goes beyond the scope of this dissertation, while there are few, if any, clear cases of philosophers who deny the realism assumption.

The second restriction is that I will not provide a full argument against dualistic II theories that deny that experienced phenomenal qualities are instantiated in the brain, such as sense data theory. The reason is that, since I cannot provide a full defense of physicalism below in Chapter 8, type physicalism is largely assumed. Since sense data theory, as I have stated it, claims that experienced phenomenal qualities belong to non-physical objects, it is committed to the falsity of physicalism. Thus, a consequence of physicalism being largely assumed is that I cannot provide a full argument

against sense data theory and other dualist II theories. As I briefly touched upon in the introduction, my argumentation for brain theory has four parts. In the first step, I argue that theories which deny the experience thesis, namely EX theories, face serious problems. I do this through two arguments: the “knowledge-acquisition argument” and the “value argument”.

In the second part, I argue that theories which deny the instantiation thesis, namely the EN and NN theories, face serious problems. I do this by looking closely at the various subtypes of theories that are included in this class of theories, arguing for each subtype that we face problems if we try to give hallucinated qualities the location that subtype of theories proposes.

In the third part of my argumentation, I argue for brain theory on the basis of the instantiation thesis, the experience thesis, type physicalism and the hallucination\* thesis. I do this mainly through the “identity argument.” The conclusion rules out EI theories, such as the view saying that hallucinated phenomenal qualities are instantiated in the brain and that phenomenal qualities experienced in veridical perceptions belong to external objects. After this part the only positions that have not been argued against are II-theories compatible with brain theory, such as internalist representationalism and, under one of their two respective interpretations, qualia theory and adverbialism.

In the fourth and final part of my argumentation I argue that brain theory does not face any significant problems. This step is crucial, since otherwise the fact that theories which deny the instantiation thesis or the experience thesis face significant problems would not give us good reasons to believe in these theses, nor, via the third step, to believe in brain theory.

## **5.2 Definitions and theses**

First, let us be precise about what we mean with the term “hallucination.” I will use the following definition, which is supposed to be as non-controversial as possible so that EX theories will not be incompatible with the terminology I will be using when discussing hallucinations and trying to argue that phenomenal qualities are experienced in them:

*Hallucinations* are events that (a): seem to be perceptions, (b) are not distally caused, in the right way, by an external object suitable for instantiating the qualities seemingly experienced in them.

Let me explain the parts in turn. The “perceptions” part is just there for convenience, to allow my term to only designate events that we typically take to be perceptions and not, say, tables or chairs.

The “seems” and “seemingly experienced” parts are there so that we can talk about hallucinations without begging the question against EX theories. The term ‘seem’ is here used in its epistemic sense, roughly standing for what we take to be the case. Perhaps it could turn out that some such events are not experiences but rather pure cognitive confabulation. In particular, when an EX theorist claims that no phenomenal qualities are experienced in ordinary hallucinations this might be taken to imply that ordinary hallucinations are not experiences. Furthermore, an EX theorist denies that phenomenal qualities are experienced in hallucinations.

The causal condition distinguishes hallucinations from veridical perceptions. I will not specify the “right way” clause further since the details of it do not matter for our purposes. Instead, I will give a couple of examples. If you have a perception as of an apple, caused by an apple seen in the ordinary way, this is an example of an external object causing a perception in the right way. On the other hand, a case where a similar experience is caused not by an apple but by a drug or an evil scientist manipulating your brain, no external object caused the experience in the right way.

The cases above also give examples of external objects which are not suitable to instantiate the phenomenal qualities experienced in an experience. Neither an evil scientist nor a drug is suitable for instantiating the various redness and roundness qualities of an apple.<sup>101</sup>

Note that by my definition many dreams are also hallucinations. To be specific, the dreams that have perceptual phenomenology and so seem to be perceptions are hallucinations.<sup>102</sup> Even if dreams perhaps would not be classified as hallucinations by ordinary linguistic norms, it clearly follows

---

<sup>101</sup> What if the evil scientist manipulates your brain by connecting it to a computer which creates a simulated world, as in the movie “The Matrix”? David Chalmers (2005) has argued that a subject connected to a computer, Matrix-style, so that she experiences a simulated world, could acquire knowledge about objects in this world. These objects would all be constituted by computational processes in the computer. Perhaps one could argue that in the described scenario such computational processes are suitable for instantiating the phenomenal qualities the subject experiences. However, whether that is the case or not will not matter for my argumentation since I will focus on hallucinations where there are no such objects that could be considered as perceived objects that are suitable for instantiating phenomenal qualities.

<sup>102</sup> Perhaps some dreams have only emotional phenomenology and no perceptual phenomenology.

given the criteria in my definition that many dreams, if not all, are hallucinations. When we have them, these dreams seem to be perceptions, but they are not caused by external objects in the right way. Thus, they count as hallucinations.

A consequence of this definition is that atypical "hallucinations" such as pathological confabulation cases and similar cases where we are unsure whether we think phenomenal qualities are experienced do not count as hallucinations. However, even if a view such as Fish's EX theory is true, on which ordinary hallucinations are very similar to such pathological confabulation cases this does not prevent ordinary hallucinations from counting as hallucinations in my sense since even on these views ordinary hallucinations *seem* to be perceptions, as (a) above demands.

To simplify things further, there are two types of non-veridical experiences I will not consider when talking about "hallucinations." These are "illusions" and "impure hallucinations." As I mentioned in Chapter 3, illusions are experiences where an external object is perceived, but it is perceived as being different from how it really is. A pure hallucination is an experience where the subject only has hallucinatory experiences at that time. Examples include dreams. An impure hallucination is one where the subject both has hallucinatory and veridical experiences at the same time. For example, someone might correctly perceive a table while also hallucinating an apple lying on the table.

Below when using the term "hallucination" I will only consider "pure" hallucinations. The reason for this restriction is that, as we shall see below, we only need conclusions about pure hallucinations for my generalizing argument in Chapter 9 to work. But if that argument works it will not just show that phenomenal qualities are instantiated in the brain in typical veridical perceptions but also that they are instantiated in the brain in impure hallucinations and illusions.

I will use "veridical perception" as a term for all perceptions that are not illusions, hallucinations in my sense, or impure hallucinations. Furthermore, from now on when I use 'hallucination' or 'veridical perception' I will use the terms in the sense I have defined here, unless otherwise is stated.

A type of hallucination, which is particularly important to my argumentation, is hallucinations where the subject is in the same global brain state as when having a certain veridical perception. I will introduce a special term for these: 'hallucinations\*'.

*Hallucinations\** are hallucinations where the subject S is in the same global brain state as when S has a certain veridical perception V.

Note that hallucinations\* are relative to subjects and veridical perceptions. For each hallucination\*, I will call the veridical perception V where the subject is in the same global brain state the “corresponding veridical perception.” Similarly, for each veridical perception I will call any hallucination\* where the subject is in the same global brain state a “corresponding hallucination\*.”

We will see some examples of hallucinations\* below when I present the super-scientist thought experiment. But before that, I will present the theses that I will argue for. In the first two parts of my argumentation I will argue for the three theses below:

*Hallucination\* thesis (H\*T):* For each veridical perception there is a corresponding indistinguishable possible hallucination\*.<sup>103</sup>

*Experience thesis (ET):* Phenomenal qualities are experienced in hallucinations.

*Instantiation thesis (IT):* All phenomenal qualities experienced in hallucinations are instantiated in existing internal objects.

Before concluding this chapter let us investigate the evidence for the hallucination\* thesis. The argument for it is simple and consists in considering the thought experiment below, which we might call “the super-scientist thought experiment,” and noting its implications.<sup>104</sup>

Take an arbitrary veridical perception E; say the one you are having at this very moment. For simplicity assume E is an experience of a stationary scene, such as a scene where you see a book lying on a table. Now imagine that the external objects which you perceive, as well as everything else in front of you, are removed while at the same time a team of super-scientists artificially keep your brain’s perceptual input processes in a constant state. This could in theory be accomplished in several different ways. For example, they could send the same type of light to your retinas as when you were seeing the thing, or they could send the same type of signals through your optical nerve. Perhaps they could even directly affect your visual cortex, say through

---

<sup>103</sup> The relevant strength of the possible modality here is nomic possibility, that is, the thesis claims that such a corresponding hallucination is compatible with the actual laws of nature.

<sup>104</sup> This thought experiment is inspired by the discussion of the causal generalizing argument in Robinson 1994 (ch. iv) and Johnston 2004.

sophisticated transcranial magnetic stimulation, and thus keep it in the same state throughout your experience. Regardless of the specific method few would deny that in such a case you would take yourself to have a perception. In fact, the only way to deny this is to accept “long-range causation,” in other words, that external objects or the lack of them could directly affect the brain without any mediating causes, such as light or sound waves. But in addition to being implausible, that goes against everything we know empirically about the brain and perceptual processes. So let us call this seeming-experience event H. As the scenario is possible the event H is also possible. Since, by assumption, the external objects were removed there are no external objects causing the seeming-experience through an ordinary perceptual process, mediated by external senses. Thus, H is a hallucination. Now, note that in the scenario described you will also be in the same global brain state as when having E, unless there is long-range causation. Thus H is a possible hallucination\* and E is a corresponding veridical perception.<sup>105</sup> Since E was chosen arbitrarily we can conclude that for each veridical perception there is a corresponding possible hallucination\*. This shows that for each veridical perception there is a corresponding possible hallucination\*.<sup>106</sup>

Furthermore, note that hallucinations\* are indistinguishable from their corresponding veridical perceptions. We can see this by again looking at the super scientist thought experiment. Unless there is “long-range causation” the subject in the thought experiment cannot distinguish between E and H, regardless of whether she is having E or H, since she is in the same brain state in both cases. As she is a normal subject that is not impaired in any way it seems reasonable to say that E and H are indistinguishable. Because the hallucination\* which corresponds with a veridical perception is indistinguishable from it, I will sometimes call this hallucination\* the “corresponding indistinguishable hallucination” of the veridical perception.<sup>107</sup> Thus, we may conclude that H\*T is true.

---

<sup>105</sup> Strictly speaking, if the described scenario would happen there would be minor differences between the global brain states in H and E. This is because in the described scenario, although the input to the brain is held constant, some small amount of time has passed between the moment the subject had E and the moment she has H. Since the whole brain is not static during this time period there have been some small changes in brain states. We could avoid these minor differences by comparing H with the veridical experience E2 the subject would have had at the same time if the super-scientists had not intervened. The global brain states of H and E2 should be the same. Thus, we can get the same conclusion and avoid this complication by using E2 instead of E as a veridical perception that corresponds to H.

<sup>106</sup> We can give the same arguments for illusions and impure hallucinations as we gave here for veridical perceptions. By using the same thought experiment and considerations we can conclude that each impure hallucination and illusion has a corresponding (pure) hallucination\*. Thus we get an analogue to H\*T for impure hallucinations and illusions.

<sup>107</sup> If there can be imperceptible differences between experiences, there might be several distinct hallucinations that are indistinguishable from a single veridical perception, V. In this case V does not have a unique indistinguishable hallucination. However, given the uncontroversial assumption that seeming experiences cannot be distinct unless there is a difference in the physical brain states of their subjects, V still has a unique corresponding *hallucination\**. In case there can be imperceptible differences between

# 6. The experience thesis

## 6.0 Introduction

In this chapter, I argue that the theories that deny the experience thesis, which claims that we experience phenomenal qualities in hallucinations, face serious problems. I show the problems using two arguments: the “knowledge-acquisition argument” and the “value argument”. Since I assumed in the last chapter that XX theories are false, the remaining theories which deny the experience thesis are EX theories. As we can recall from Chapter 4, the most plausible version of EX theories claims that, when hallucinating, although we do not experience phenomenal qualities, we are cognitively related to phenomenal qualities, in virtue of having beliefs about them. This can be expressed by saying that according to such EX theories we cognize phenomenal qualities in hallucinations, rather than experience them.

## 6.1 The knowledge-acquisition argument

In this section I give my first argument for the experience thesis (ET), the thesis that phenomenal qualities are experienced in hallucinations. This is the knowledge-acquisition argument. After briefly describing the core idea behind it I will put it more formally in terms of premises and conclusion before looking in turn at how well supported the premises are.

We might express the knowledge-acquisition argument as follows:

P1 We can acquire knowledge of what phenomenal qualities are like through hallucinations.

P2 The best explanation of the fact that we can acquire knowledge of what phenomenal qualities are like in hallucinations is that we experience phenomenal qualities in hallucinations.

C1 (From P1 and P2 by inference to the best explanation) We experience phenomenal qualities in hallucinations (ET).

---

experiences, we can call this hallucination\* the *corresponding indistinguishable hallucination* of V, as hallucinations\* are indistinguishable from their corresponding veridical perceptions.

The argument is not deductively valid, but as long as we accept inference to the best explanation as a good inference procedure the conclusion C1 follows from P1 and P2

Let us first look at the case for P1. Mark Johnston has made a similar argument against disjunctivism in Johnston (2004). There he gives several examples of ways in which we can gain knowledge about what phenomenal qualities are like from hallucinations:

I can secure my first singular reference to the quality cherry red or to the structural property C major by way of hallucinating a scene or a tune. Frank Jackson's Mary could come to know what red is like by hallucinating a red thing or by having a red afterimage. Indeed, we shall later encounter a case which implies that, as a matter of empirical fact, the paradigm red – the reddest of reds – can only be presented in delusive experience. One can come to know what "supersaturated" red is like only by afterimagining it. While one is imagining it, one could compare how much more saturated it is than the reds exhibited by the reddest of the standard Munsat color chips, there before one on the table. Likewise, a painter might discover in hallucination strange, alluring color, which he then produces samples of by mixing paints in a novel way. Here we have all the signs of *de re* knowledge of quality. One comes to know what certain qualities are like, and so one is able to place them in a quality-space with other qualities of the same family.<sup>108</sup>

Inspired by his argumentation we can give three parallel examples, in support of P1. First, consider Mary, who was raised in a black-and-white environment in Jackson's knowledge argument. She could get knowledge of what it is like to experience phenomenal red through a hallucination or afterimage.<sup>109</sup> Second, even normal people can acquire knowledge of what phenomenal qualities are like through hallucinations. Consider the color supersaturated red. The typical way of learning what it is like is through an afterimage, a type of hallucination.<sup>110</sup> Third, consider the case of an artist who discovers a new interesting paint color in a dream and then mixes it afterwards. It seems the artist got knowledge about what the phenomenal quality was like through having the dream.

Note that there is nothing special about color qualities. What goes for color in the examples above should apply to sound qualities or just about any type

---

<sup>108</sup> Johnston (2004, 130-131).

<sup>109</sup> The knowledge argument will be discussed in more detail below in Chapter 8.

<sup>110</sup> Note that an afterimage is typically only an impure hallucination. However, it seems that if we can learn what phenomenal qualities are like from the hallucinatory parts of impure hallucinations this gives us reason to believe we could learn what phenomenal qualities are like also from pure hallucinations.

of phenomenal quality. Thus, we seem to have a decent prima facie case for P1.

What about P2? That we experience phenomenal qualities in hallucinations provides an explanation of why we can acquire knowledge of phenomenal qualities in hallucinations. We get the knowledge about the qualities by experiencing them, just as we get knowledge about phenomenal qualities when experiencing them in veridical perceptions.

An EX theorist would deny that we get knowledge of what it is like to experience phenomenal qualities in this way, and instead claim that we can get knowledge of what qualities are like through hallucinations without experiencing them. But just denying that one explanation is true does not provide an explanation, much less the best explanation.

One possibility is to somehow invoke the fact that the downstream cognitive effects of hallucinations are the same as for corresponding veridical experiences. However, to rival the explanation that can be provided by a theorist who accepts the experience thesis, a complete account will have to be worked out and made plausible, something which has not been done yet.<sup>111</sup>

## **6.2 The value argument**

In this section I give my second argument for the experience thesis. This is the value argument. After briefly describing the core idea behind it and presenting some concepts that are important for understanding the argument I will put it more formally in terms of premises and conclusions. Then, finally, I will look at the premises, two objections and whether the conclusions follow from the premises.

---

<sup>111</sup> For example, Fish (2009, 2010) does not consider this problem and Johnston (2004) himself does not discuss if and how the disjunctivist could respond by giving such an account. Pautz (2007) discusses a version of Johnston's epistemological argument, with the specific conclusion that we are aware of uninstantiated properties in hallucinations, since we can get knowledge about them from hallucinations. There he sketches an alternative account of how we get such knowledge without direct awareness of uninstantiated properties. However, on his alternative account we get the knowledge through "sensorily entertaining a proposition." That is a representational relation which is stipulated to be identical with the property of having an experience with a certain phenomenal character. Thus that kind of alternative account cannot help an EX theory, since the EX theory denies that hallucinations have phenomenal character, and so cannot claim that we sensorily entertain any propositions when having a hallucination.

Very briefly, I will argue as follows. Many experiences are valuable because of the qualities experienced in them. Among the clearest examples are pains and pleasures. But it seems like corresponding indistinguishable hallucinations also would have similar value. The value of these hallucinations depends on phenomenal qualities being experienced in them, so it follows that phenomenal qualities are experienced in many hallucinations. But then it seems reasonable to conclude that phenomenal qualities are generally experienced in hallucinations.

To construct a more precise value argument from this chain of thinking, let us introduce a new concept:

*Quality-dependent value* is value that something has only if certain phenomenal qualities are experienced by its subject. The value is said to *depend* on these qualities being experienced.

To make the presentation simpler, below in this chapter, instead of saying that “value depends on phenomenal qualities being experienced”, I will use the shorter expression that “value depends on phenomenal qualities”.

Note that the phenomenal qualities in question can be different qualities in different experiences.

Another concept that is important to the argument is non-instrumental value:

*Non-instrumental value* is value that is not instrumental.

A thing with non-instrumental is valuable for its own sake, or as one might express it, “is an end in itself.” This means that to account for its value we do not have to refer to any other thing for which’s sake it is valuable.<sup>112</sup> It is important to note that non-instrumental value need not be intrinsic value, in the sense that the valuable object has its value only in virtue of intrinsic properties. That something has non-instrumental value only implies that it has value that is not instrumental to achieving something else<sup>113</sup>. For example, you might hold that the object that is your first guitar has non-

---

<sup>112</sup> On non-instrumental value, see Samuelsson (2008, 72) and O’Neill (1992, 119).

<sup>113</sup> John O’Neill distinguishes between three senses of ‘intrinsic value’. First, ‘intrinsic value’ might mean non-instrumental value as I have described it above. Second, it might mean value objects have in virtue of “intrinsic properties” as in non-relational properties. On a weaker interpretation of “non-relational” these non-relational properties are properties of an object that can persist regardless of the existence of other objects. In a stronger interpretation the non-relational properties are properties of an object that can be characterized without reference to other objects. Third, ‘intrinsic value’ might mean objective value, that is value that is had by an object independently of the valuations of valuers. See O’Neill (1992, 119-120).

instrumental value, even if this value depends on the fact that it was your first guitar and thus is non-intrinsic value.<sup>114</sup>

We can now express the argument more formally as follows:

P1 Many hallucinations have non-instrumental value

P2. If a hallucination has non-instrumental value then it has quality-dependent value.

C1 (By P1, P2) Many hallucinations have quality-dependent value

C2 (By C1 and definition of quality-dependent value) Phenomenal qualities are experienced in many hallucinations.

C3 (reasonable conclusion from C2) We experience phenomenal qualities in all hallucinations (ET)

The argument, up to C2, is deductive and valid but the last step, the inference to C3, is not deductive. However, as we shall see below, it is a reasonable inference. But before that, let us look at the premises, in turn.

### 6.3 The first premise

The first premise says that many hallucinations have non-instrumental value. How many? Well, at least as many as the veridical perceptions that have non-instrumental value. This follows since it seems that if a veridical perception has non-instrumental value then an indistinguishable hallucination would also have it. Below, I will present two specific examples of hallucinations that have non-instrumental value, positive or negative: pain hallucinations and positive hallucinations.

---

<sup>114</sup> This guitar example was inspired by a slightly different guitar example in Samuelsson (2008, 188-189) where it is argued that a person's first guitar can have non-instrumental value in virtue of having been massively instrumentally valuable to that person. Other supposed examples of objects with non-instrumental non-intrinsic value include Lincoln's pen (Kagan 1998, 285) and the dress of Princess Diana (Rabinowicz, and Rønnow-Rasmussen 1999, 41). Arguably, these can have non-instrumental value in virtue of their historical connections to the individuals Lincoln and Diana and famous historical events including them. However, they have this value in virtue of the relational properties mentioned above, rather than any intrinsic properties of the pen and dress in question. Note that I here refer to "relational properties" in the stronger sense described in the note above.

First, consider pain hallucinations. Humans can have several different types of pain hallucinations and illusions. One of these is phantom limb pain. After losing a limb it is common to experience chronic pains which are felt where the lost limb would be. These are phantom pains. The types of sensations and the severity of the pains vary from person to person but for some it can be extremely agonizing.

Are phantom pains bad? In many cases they are severe and debilitating. We clearly *think* they are bad and feel sorry for their victims. Few would deny that they are bad, except those who deny that there is something like badness in general. But as we will see below, a modified version of the value argument can give us the same conclusions even if there is no such thing as value. Furthermore, at least part of their badness seems to be non-instrumental. Phantom pains are not just bad because they are debilitating: having the experience is also bad in itself.

Second, consider positive hallucinations. Cases include dreams, in particular, lucid dreams, transcendental experiences and drug experiences. We consider many of these experiences valuable. If the phantom pain experiences discussed above are valuable then these should also count as valuable.

The clearest and most versatile examples of positive hallucinations are provided by the so-called “experience machine”. It is a hypothesized machine which when we connect ourselves to it produces any types of experiences that we wish to have; and in particular it can produce any pleasure. For example, we can experience any type of pleasant aesthetic visual experience with various intricate color and shape qualities, any type of adventure or life-story.

Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life’s experiences? If you are worried about missing out on desirable experiences, we can suppose that business enterprises have researched thoroughly the lives of many others. You can pick and choose from their large library or smorgasbord of such experiences, selecting your life’s experiences for, say, the next two years. After two years have passed, you will have ten minutes or ten hours out of the tank, to selection the experiences of your *next* two years. Of course, while in the tank you won’t know that you’re there; you’ll think it’s all actually happening. Others can also plug in to have the experiences they want so there’s no need to stay unplugged to serve them...

Would you plug in? *What else can matter to us, other than how our lives feel from inside?* Nor should you refrain because of the few moments of distress between the moment you've decided and moment you're plugged. What's a few moments of distress compared to a lifetime of bliss (if that's what you choose), and why feel any distress at all if your decision is the best one (Nozick 2005, 311)?

Nozick uses the experience machine thought experiment to argue against hedonism. But note what he argues against is not the claim that experiences have value. What he argues against is the much stronger claim that nothing but experiences has value. That such a strong hypothesis is even discussed indicates that we think experiences, including hallucinations, can at least have some non-instrumental value. More generally, note that P1 is not incompatible with the idea that "false pleasures," pleasures based on false beliefs, might be worth less than indistinguishable pleasures based on true beliefs.<sup>115</sup>

It seems clear that if the corresponding indistinguishable "veridical" pleasure is good then the pleasure given by the experience machine will also be good. Perhaps it is not as good as when experienced outside the machine, but it is still better than nothing and so it has some value.<sup>116</sup> We can see this more clearly if we compare a pleasant stay in the experience machine with a stay in the machine where nothing is experienced. Having these pleasant experience-machine-experiences is better than experiencing nothing at all, which is all we need for the value argument.

One might wonder if pleasures can be hallucinatory at all. It does sound a little bit strange to talk about "pleasure hallucinations". However, note that we are not concerned with the ordinary term 'hallucination' here, we are concerned with my stipulated technical term 'hallucination' and all that requires is events that (a): seem to be perceptions, and (b) are not distally caused, in the right way, by an external object suitable for instantiating the qualities seemingly experienced in them. Both (a) and (b) are fulfilled for the experiences we are considering here. Thus, they are hallucinations in the relevant sense.

Before finishing the investigation of P1 let me note that under some assumptions, denying this premise has some very radical implications. As

---

<sup>115</sup> For example, the pleasure someone takes in her good relationship with her partner would count as such a false pleasure if the relationship were not good; say because the partner was cheating on her without her knowing it.

<sup>116</sup> The argument works even if the value of non-hallucinatory positive experiences lexically dominates the value of positive hallucinatory experiences, so that no amount of hallucinatory experiences could be worth as much as the least valuable positive veridical experience.

Bostrom (2003) argues, there are empirical reasons to give some credence to the hypothesis that we are living in a simulation. In that case, if P1 is false all our experiences would lack non-instrumental value. Furthermore, to the extent that instrumental value is instrumental for quality-dependent value, all of our experiences would lack any value.

## 6.4 The second premise

Now let us move on the second premise, P2, which says that if a hallucination has non-instrumental value, then it has quality-dependent value.

Note that when I say that hallucinations have quality-dependent value, this only implies that *at least a part* of their value depends on their qualities being experienced. P2 does not imply that they have value *only* because of their qualities. Thus P2 is not incompatible with views which hold that pains and pleasures can be valuable for other reasons or instrumentally valuable.

I will first look at why pain hallucinations have quality-dependent value, and then look at why positive hallucinations have quality-dependent value.

There are at least two reasons to believe that the badness of pain experiences in general and pain hallucinations in particular depends on their qualities. First, it seems that way when we attentively introspect pains and consider what we dislike about them. When having a pain, perhaps from pinching your arm, introspect the pain. What is it about the experience that you dislike? What is bad about it? When you try to introspect it I predict that you find certain annoying qualities; pain qualities. Your dislike seems to be directed towards these pain qualities. Thus, introspection gives us reason to believe that pain qualities contribute to the badness of pains.<sup>117</sup>

---

<sup>117</sup> In addition, the pain theorists who explicitly mention qualities tend to reach a similar conclusion: that the qualities are the target of our negative attitudes to pains. See for example, Rachels (2000), Tye (2000) and Broad (1930). For the analogue claim about pleasures see Alston (1967). As it seems plausible that what is bad about pains is what our negative attitudes are directed at I conclude that this also give us prima facie reasons to believe P2 in the case of pain. We can make a similar distinction for types of accounts of pleasures as for pains. It is common to distinguish between two general views on pleasure: distinctive feelings views and attitudinal views. The former says what all pleasures have in common is a distinctive feeling. The latter says that what they have in common is that they are experiences which are targeted by a certain attitude. See Brax (2009, 20). Among others, this distinction has been made by Feldman (1997), Gosling (1969) and Crisp (2006).

Second, for a certain pain experience consider the counterfactual case where you are a partial zombie and have an experience which is the same in all respects as the original pain experience, but which does not include any phenomenal pain qualities. In this counterfactual case, the experience would not have been as bad as it was with the pain qualities. The value that depends on the qualities in it would be missing. To put the same point differently, we would be satisfied with a pain-killer that did nothing more than prevent us from experiencing pain qualities. Since the removal of experiences of pain qualities makes pains less bad, it follows that part of the negative value of pains, including pain hallucinations, depends on phenomenal qualities.

Does part of the non-instrumental value of pleasures depend on phenomenal qualities? This seems hard to deny since we can give analogues of the two arguments for pains above.

First, if we introspect a pleasurable experience, we see that we value the qualities in the experience. To give one example, in an aesthetic experience, the experience of color and shape qualities seem to contribute to the value. The same also goes for lucid dreams and transcendental experiences. I, myself, recall a lucid dream that made a great impression on me and seemed like a very valuable experience. It was an experience as of colors and shapes that seemed to be clearer, more vivid and to have finer details than things perceived in normal veridical perceptions. Now, looking back at it, it seems like at least part, if not all, of the value depended on the qualities being experienced.

Second, if we counterfactually consider the case where a certain pleasure experience would occur without the experience of any qualities, this result seems even harder to deny. What would an aesthetic experience of a beautiful piece of art be worth without the experiences of the respective color qualities? What would the experience of one's favorite music be worth if one did not experience any sound qualities? Although it might seem plausible that many pleasures are mostly, or even only, valuable because of the phenomenal qualities we experience in them note that all we need here is the much weaker claim that the qualities contribute to their value so that these experiences would be *less valuable* if no phenomenal qualities were experienced in them.

## 6.5 The no value objection

What if there are no values? Some philosophers are anti-realists about values and claim that judgments about things being valuable are either false or meaningless.<sup>118</sup> But even if anti-realism about values is true, it might be possible to reconstruct a working version of the value argument. Consider phantom pains. Instead of their badness, we could focus on the fact that subjects of phantom pains have their preferences frustrated; preferences which are directed at the absence of pain qualities. People strongly want to avoid experiencing severe pain qualities; they have what we might call "no-pain-preferences."

An analogue of the first premise then says that we have preferences against phantom pains that are frustrated. An analogue of the second premise says that some of these frustrated preferences against phantom pains are no-pain-preferences.

We can reason in the same way for positive hallucinations.

Below I will give a short sketch of how a modified argument might proceed, after introducing the term quality dependent preference:

*A Quality-dependent preference* is a preference that is directed for or against phenomenal qualities and so is only frustrated or satisfied if certain phenomenal qualities are experienced by its subject. The preference is said to *depend* on these qualities.

We can now express *the preference argument*:

P1 We have preferences for many hallucinations, and these preferences are sometimes satisfied. Similarly, we have preferences against many hallucinations, and these preferences are sometimes frustrated.

P2. For each of these hallucinations some of these preferences are quality-dependent.

C1 (By P1, P2) For many hallucinations there are quality-dependent preferences that are frustrated or satisfied.

---

<sup>118</sup> Those who claim that moral judgments are systematically false are called error theorists; see Mackie (1977) for the classical exposition. Those who claim that moral judgments lack cognitive meaning on the other hand are called non-cognitivists.

C2 (By C1 and definition of quality-dependent preference)  
Phenomenal qualities are experienced in many hallucinations.

C3 (reasonable conclusion from C2) We experience phenomenal  
qualities in hallucinations (ET)

Here, the premises P1 and P2 and the conclusion from C2 and C3 in the preference argument can be supported in the same ways as P1 and P2 in the value argument.

## 6.6 The disassociation objection

One possible objection to P2 is that pain, and thus the pain qualities, can be disassociated from the badness of pain in certain empirical cases, and that this indicates, contrary to P2, that the badness of pains does not depend on phenomenal qualities. If pain qualities can be experienced in experiences that are not bad, then it seems these qualities are at least not intrinsically bad. That might give us reason to think they are not bad at all. I will argue below that such disassociation cases do not pose a problem for the value argument. But before that, let us first take a closer look at the disassociation cases.

Several cases of pain without unpleasantness and even a few cases of alleged pain unpleasantness without pain have been reported. Examples of the first type of disassociation include prefrontal lobotomy, pain under morphine induced analgesia and pain asymbolia.<sup>119</sup> In such cases the patients, when given pain stimuli, can identify pain sensations but are not bothered at all by the pains; according to their reports the sensations do not feel the least bit unpleasant to them.<sup>120</sup> Clear cases of the second type of disassociation, pain unpleasantness without pain, are much rarer; what may well be the only confirmed case is described below:<sup>121</sup>

[At higher intensities of cutaneous laser stimulation] the patient spontaneously described a "clearly unpleasant" intensity-dependent feeling emerging from an ill-localized area "somewhere between fingertips and shoulder" that he wanted

---

<sup>119</sup> Grahek (2007) argues that pain asymbolia is the only genuine case of disassociation, as conceived in the philosophical literature.

<sup>120</sup> For more on these cases see Aydede (2013). Grahek (2007) offers what is perhaps the most thorough philosophical investigation available.

<sup>121</sup> According to Aydede (2013) this is the only uncontroversial case of pain affect without pain sensation. For an older possible description of such a case, see Head and Holmes (1911).

to avoid. The fully cooperative and eloquent patient was completely unable to further describe quality, localization, and intensity of the perceived stimulus. Suggestions from a given list containing "warm," "hot," "cold," "touch," "burning," "pinprick-like," "slight pain," "moderate pain," and "intense pain" were denied (Ploner et al., 1999, 213 cited in Aydede 2013).

Pain scientists and philosophers have reacted somewhat differently to disassociation cases. A common reaction among philosophers is to see these cases as a reason for adopting a so called attitudinal view of the badness of pains. See for example, Clark (2005), Parfit (1984, 501), Brandt (1998, 37-38) and Hall (1989). Pain scientists, on the other hand, have developed a view according to which pains have three different components.<sup>122</sup> In addition to the familiar sensory component of pain, there is also an affective component and a cognitive component. In this model, the cases of pain without unpleasantness are interpreted as pains where the sensory component is intact but the affective component is lacking. Cases of unpleasantness without pain, on the other hand, are interpreted as pains where the sensory component is missing but the affective component is intact.

This model opens up the possibility for what we might call a *complex account of the badness of pains*, the position that pains have two qualitative components, and that only one of these components, the affective pain qualities, is bad in pains. Since according to the model these affective pain qualities are absent in cases of pain without unpleasantness, they explain why such experiences are not bad. On the other hand, since affective pain qualities are experienced in ordinary pains, they can still explain why ordinary pains are bad.

Now we can see that both of these two models of the badness of pains, which people have thought are supported by disassociation cases, are compatible with P2. On attitudinal theories, disassociation cases are compatible with P2. Pains and their qualities are bad because we hold a certain negative attitude towards them. Cases of pain without unpleasantness are explained as situations where we experience the same qualities but do not hold the negative attitude towards them. The pain qualities are bad in typical cases of pain although the very same qualities are not bad in disassociation cases. Thus, typical non-disassociation pain experiences have quality-dependent value.

---

<sup>122</sup> Melzack and Casey (1968) was the groundbreaking paper that first proposed the three component model of pain.

On a complex view of the badness of pains, disassociation cases are compatible with P2. Although pains are not bad because of *sensory* pain qualities, they are bad because of other qualities, *affective* pain qualities.<sup>123</sup> Thus, typical non-disassociation pain experiences have quality-dependent value.

Still, even if P2 is compatible with a complex view of the badness of pains, one worry might remain. Although affective pain qualities are phenomenal qualities in the broad sense I defined in Chapter 2, one might think that they are so different from typical sensory qualities, such as phenomenal redness qualities or sensory pain qualities, that we should distinguish between these two different kinds of qualities. Thus, it might seem possible to accept that we experience affective pain qualities when having hallucinations while at the same time denying that we experience typical sensory qualities in hallucinations. Although the value argument up to C2 would then be correct for pains, we would not have good reasons to accept to inference from C2 to C3.

Even if this worry is justified, it is no problem for our wider argumentation. As we shall see below the inference from C2 to C3 mostly depends on pleasures, and pleasures are importantly different in the relevant respect. For pleasure, there does not seem to be similar disassociation cases where an affective “hedonic tone” component is separated from a sensory component of the pleasure. This might be explained by an asymmetry between pains and pleasures. For pains, it can seem reasonable that only one component of the pain has (negative) value. For typical pleasures and other positive experiences, on the other hand, it is commonly believed that their value cannot be reduced to the value of a single common component, a “hedonic tone.”

## 6.7 The final inference

Let us now look at the final inference in the argument, the conclusion of C3 from C2. This inference seems justified, for the following three reasons:

R1: Phenomenal qualities are experienced in *many* hallucinations, according to C2, namely those that have non-instrumental value.

---

<sup>123</sup> On first glance one might think affective pain qualities seem too far removed from the paradigmatic phenomenal qualities we used to define the class. However, on reflection they are not stranger than “feeling phenomenal qualities” or emotional phenomenal qualities, such as sadness or anger qualities.

R2: A wide range of phenomenal qualities are experienced in these hallucinations.

R3: There is no particular reason to believe that phenomenal qualities should not be experienced in hallucinations which lack non-instrumental value.

Before concluding the argument and the chapter, let us briefly look at these three reasons, in turn. First, concerning R1, we might ask precisely how many types of experiences have non-instrumental value. As was noted above, due to the possibility of indistinguishable hallucinations, these should at least be about as many as the number of veridical experiences that have non-instrumental value.<sup>124</sup>

Many different experiences can be pleasurable. Just to mention a few, there are dreams, emotions, feelings and various sensory pleasures, such as sights, sounds, tastes and intimate activities. More complex aesthetic pleasures include experiences of plays, poetry, music and art. Experiences of nature, or other beautiful things, may also count as aesthetic pleasures in some sense.

Note that I do not use the term ‘pleasure’ in a narrow way, as it is perhaps sometimes used, just to denote simple sensory pleasures like eating or having sex. Rather, I am using the term ‘pleasure’ in a broad sense to include all kinds of positive experiences. In addition to the non-veridical pleasures that are indistinguishable from veridical perceptions there is also a wide range of positive non-veridical experiences that are distinguishable from veridical perceptions, including dreams, drug-experiences and transcendental experiences.

Pains and other negative experiences may not be as varied as positive experiences, but they do also contribute to increasing the range of hallucinations with non-instrumental value. In addition to phantom pains, examples here include experiences such as nausea, nightmares and bad drug trips.

Second, concerning R2, there is a huge range of types of qualities that hallucinations with which non-instrumental value depend on. Pleasurable experiences, such as the ones mentioned above, seem to be able to depend on just about any phenomenal quality and, due to people’s varying tastes, a wide range of combinations of phenomenal qualities can be pleasurable at

---

<sup>124</sup> In some cases a veridical experience might be thought to be valuable partly because it is a veridical perception. For example, perhaps some would consider a veridical perception of a certain famous person, say Elvis, to be more valuable than an indistinguishable hallucination.

least to someone. The exceptions would be qualities, or patterns of qualities, typically associated with unpleasant experiences.<sup>125</sup> But these, on the other hand, are for that reason, precisely those qualities that we have reason to believe may determine the non-instrumental value of negative experiences.

Third and finally, there is no particular reason to believe that phenomenal qualities should not be experienced in hallucinations which lack non-instrumental value. We have reason to believe that qualities are experienced in hallucinations that have non-instrumental value. However, no reason has ever been mentioned why no phenomenal qualities should be experienced in hallucinations that are neutral.

To conclude, if a wide range of phenomenal qualities are experienced in a wide range of hallucinations, and we have no particular reason to believe that no phenomenal qualities are experienced in hallucinations outside this range we should conclude that phenomenal qualities are experienced in hallucinations. Thus, together with R1 and R2, R3 gives us a good reason to believe C3, given C2.

---

<sup>125</sup> However, there are some cases which might seem like pleasant experiences with some unpleasant components. The experiences of those who enjoy very spicy food, horror movies or sado-masochism seem to be examples.



# 7. The instantiation thesis

## 7.0 Introduction

In the last chapter I argued that EX theories are false and that ET, the thesis that phenomenal qualities are experienced in hallucinations, is true. In this chapter I will argue that theories that deny the instantiation thesis have serious problems. These are EN and NN theories of phenomenal qualities, the theories which agree that we experience phenomenal qualities in hallucinations, but at the same time claim that these experienced phenomenal qualities are uninstantiated. Such theories of the phenomenal qualities are associated with representationalist theories of perception, but can also be used by disjunctivist theories of perception. As I mentioned in chapter 4, this class of theories has three subcategories. First, there is Property-ism, which claims that phenomenal qualities experienced in hallucinations are uninstantiated properties themselves. Second, there is Nothingism, which claims that phenomenal qualities in hallucinations are experienced but do not exist, simpliciter. Third, there is Meinongianism, which claims that phenomenal qualities in hallucinations are non-existing entities which "have being." I will go through these three subcategories in turn, arguing that each has serious problems. Most of the chapter will be spent investigating, and arguing against, Property-ism.

## 7.1 Property-ism

The Property-ist claims that the phenomenal qualities experienced in hallucinations are uninstantiated properties themselves. However, this claim is not specific enough for us to be able to evaluate it properly, since the claim will give the qualities different locations depending on what the correct metaphysics of properties is. So before we look closer at Property-ism, let us briefly overview the main theories of the nature of properties.

Many philosophers take properties, as well as relations and kinds, to be universals.<sup>126</sup> The defining feature of universals is that they can be

---

<sup>126</sup>In addition to properties proper there might also be entities such as "species" and "natural kinds". Examples include the species tiger and the kind gold. As with properties there are different theories about species and natural kinds. For example, these entities might be universals, or non-universal abstract objects such as sets. Below, both universals and abstract objects will be considered as possible locations for phenomenal qualities. Since neither kinds nor species thus introduce any new alternatives for the location of phenomenal qualities I will not consider them separately from universals.

instantiated by many things. There are two main theories about the nature of universals: Aristotelianism (immanent realism) and Platonism (transcendent realism). However, some philosophers, nominalists, claim that there are no universals, and thus cannot say that properties are universals. We can distinguish three main nominalist theories of properties: “eliminative nominalism”, “reductive nominalism” and trope theory.<sup>127</sup>

Together, the two theories of universals and the three versions of nominalism give rise to five different theories of the metaphysics of properties. In the sections below, I will briefly describe these theories, and for each theory define a position consisting of the combination of that theory of properties and Property-ism, giving us five combined positions. When doing that, I will also look at which problems the combined positions face.

## 7.2 Platonism

*Platonism* (transcendent realism) says that properties are universals and that universals are abstract objects. The universals are distinct from and independent of their instances. For example, in addition to all brown things, there is also the distinct abstract object Brownness. When an object is brown it exemplifies Brownness. Even if there are no brown things Brownness can still exist.<sup>128</sup>

Abstract objects are generally conceived of as objects that are non-spatiotemporal and causally inert. Sometimes they are even *defined* as having these properties.<sup>129</sup> They exist outside time and space and unlike

---

<sup>127</sup>There is also a third type of theory that rejects universals: conceptualism. According to this position there are no universals, and words that could refer to universals, such as “redness”, instead refer to concepts, understood as mind-dependent entities. If conceptualism could somehow be combined with Property-ism, to account for phenomenal qualities experienced in hallucinations, it seems that the result would not be incompatible with brain theory, as phenomenal qualities would be mind-dependent entities, and thus reasonably located in the mind.

<sup>128</sup>What about a world in which no object is ever brown? It is possible to hold a view according to which abstract objects that are universals do not exist in worlds where they have had no instances, and only come into existence when an object instantiates them. For simplicity I will assume that if there are abstract universals they always exist, regardless of whether they have had instances or not. If anything, this assumption should put the Property-ist in a better position to provide a location for phenomenal qualities experienced in hallucinations than if the existence of abstract universals depends on their instances. The reason for this is that in a case where someone has a hallucination of, for example, phenomenal red, and that property has never existed; there is no abstract object universal of phenomenal red that can provide a location to the experienced quality. This problem will be discussed in more detail below, when I examine the combination of Aristotelianism and Property-ism.

<sup>129</sup>See Rodriguez-Pereyra (2011), Rosen (2012) and Lowe (2002) on this.

concrete objects, never take part in causal chains. This causal inertness is sometimes inferred from their being outside time and space; it is thought that causal efficacy needs spatiality.<sup>130</sup> Abstract objects need not be universals, they can also be particulars. For example, sets are sometimes taken to be particular abstract objects. Other examples are classes, propositions and mathematical objects such as numbers.

We can define the combination of Platonism and Property-ism as follows:

*Property-ist-Platonism:* Platonism is true and phenomenal qualities experienced in hallucinations are uninstantiated properties.<sup>131</sup>

The Property-ist says that phenomenal qualities experienced in hallucinations are “uninstantiated properties.” I take it that saying that a property is uninstantiated, in this context, means that the property is not instantiated in an object that is perceived by the subject that experiences the phenomenal qualities in the hallucination.

On Platonism there is nothing strange about uninstantiated properties, they are abstract objects, just like instantiated properties. The only difference is that the uninstantiated properties happen to lack instances at a certain time and place. The universals might even lack instances at all times and places but still exist.

Let us proceed to the evaluation of the combined position. Property-ist Platonism has some counterintuitive consequences when it comes to hallucinations. According to it, whenever we have a hallucination, or dream, all the qualities we experience, the colors, the shapes and the sounds, are all eternal objects that are located outside time and space. However, although such implications are counter-intuitive they do not imply that the view is false, so let us here just note that Property-ist Platonism is counter-intuitive, without holding it against the view.

Below, I will present two objections against Property-ist Platonism. The first, and main, objection against this position is the *particularity objection*. This objection says that there is a particularity to the phenomenology of hallucinations that cannot be well explained by a theory that claims that we only experience abstract universals. But first I will explain the thought behind this objection more carefully, and then look at what speaks in favor of

---

<sup>130</sup>See Rodriguez-Pereyra (2011).

<sup>131</sup>Forrest (2005) is an example of a Property-ist-Platonist.

there being a particularity to the phenomenology of hallucinations. We seem to experience properties when having hallucinations. This is one motivation for Property-ism. However, the phrase “experience properties” is ambiguous between two meanings of “properties”. First, it might mean that we experience the very properties themselves, as in that we experience the universals themselves. Second, it might mean that we experience *instantiations* of properties. What is an “instantiation”? As we will discuss in more detail when investigating trope theory, below in section 7.5, particulars which instantiate properties can be said to have *tropes* - entities that are “bits” or parts of properties. For example, each red thing can be said to have some redness in it, a certain bit of the property redness. Note that we can say that there are tropes in this sense without being committed to the claim that properties are fully reducible to tropes. I shall reserve the term ‘trope theory’ for the latter claim, which I will return to in section 7.5 below. On other theories of properties tropes might even be the reducible entities. For example, they might be complexes of properties, individuals and “exemplification nexuses”.<sup>132</sup> To emphasize that trope theory about properties is not presupposed by the particularity objection, below I will use the term “instantiation” instead of the term “trope” to refer to these bits of properties.

When we have an experience, even if it is a hallucination, phenomenologically, what we seem to experience is something more particular than just a pure universal. Having made the distinction between properties as universals and properties as instantiations, we can express this phenomenological observation more clearly as follows: when having a hallucination we seem to experience properties in the second sense of the phrase, properties as instantiations, rather than properties as universals.<sup>133</sup> This is the particularity of the phenomenology of hallucinations.

That there is a particularity to hallucinations can be supported as follows. We have a phenomenological intuition that there is a particularity to experience, in the sense that we experience instantiations, even in hallucinations. Brad Thompson expresses this intuition, although briefly:

Conscious experiences, including hallucinations, have a particularity to them that seems to rule out the position that phenomenal properties are uninstantiated properties. Whatever phenomenal properties are, they need to

---

<sup>132</sup> On the nature of tropes under trope theories of properties and under other theories of properties, see Bacon (2011).

<sup>133</sup> Of course, I do not deny that we might experience universals *through* experiencing instantiations. But at the very least, we do not *only* experience universals. Either we merely experience instantiations, or we experience both instantiations and universals

be instantiated in order to do the metaphysical job of constituting phenomenal character (Thompson, 2008, 19).

According to Thompson, the phenomenal qualities we experience cannot be “uninstantiated properties”, which means, as I interpret him, that the experienced qualities cannot be pure universals. Rather, what we experience must be *instantiations* of these qualities. Thus, the qualities have to be instantiated.

We can see this intuition especially clearly if we look at a case of a veridical perception of two qualitatively identical red apples, Lefty and Righty, experienced as being next to each other. Both are experienced as having the same specific shade of phenomenal redness; say the shade “phenomenal redness<sub>59</sub>”. In this case, it seems we do not just experience the same specific type of phenomenal redness twice, but also it seems we experience two distinct *instantiations* of phenomenal redness<sub>59</sub>: one redness that seems to reside in Lefty and a numerically distinct, although perfectly qualitatively similar, redness that seems to reside in Righty. These two “rednesses” are two instantiations of phenomenal redness<sub>59</sub>. In a case like this it seems hard to deny that if we can be said to experience redness<sub>59</sub> we can also be said to experience the two distinct instantiations of redness<sub>59</sub>.<sup>134</sup> But now, assume that the experience of Lefty and Righty was merely a hallucination. We still experience two distinct instantiations of redness<sub>59</sub>.

For the Property-ist Platonist, uninstantiated properties are pure universals, rather than instantiations, and universals are not particulars. So, since Property-ist Platonism claims that we only experience uninstantiated properties in hallucinations, it implies that we do not experience any particulars in hallucinations. But, without experience of instantiations through particulars it seems hard to account for the particularity of experience. Furthermore, Property-ist Platonism has a hard time accounting for the phenomenology of a hallucination where we experience several distinct instantiations, such as the hallucination of Lefty and Righty above. We experience two distinct instantiations of redness<sub>59</sub> here, but there is only one universal redness<sub>59</sub>. According to Property-ist Platonism these experienced redness<sub>59</sub> qualities should both be identical to that single universal. But it is hard to comprehend how that is possible, since they are

---

<sup>134</sup> Another reason there is a particularity to experience is that when we experience properties we typically experience them through their instantiations. For example, when we experience redness, we experience it through experiencing an instantiation of redness. If we can be said to experience redness itself, we experience it through experiencing the instantiation. At least, this is what happens in veridical perceptions. We have no reason to believe that things should be different in hallucinations. Thus, when we experience qualities in hallucinations, we should also experience instantiations of these properties.

two and the universal is one. Thus, the particularity objection provides a reason to reject Property-ist Platonism.

The second objection against Property-ist Platonism is the *quality causation objection*. Roughly, it says that Property-ist Platonism fails to account for the causal efficacy of phenomenal qualities experienced in hallucinations, and that this is a problem for the position. After explaining why hallucinated qualities cannot be causally efficacious on Property-ist Platonism, I will present two reasons why this is problematic.

According to Property-ist Platonism, phenomenal qualities experienced in hallucinations are abstract objects. As was mentioned above, abstract objects are causally inert. Thus, it follows straightforwardly that hallucinated phenomenal qualities cannot be causally efficacious. The Platonist might argue that universals, despite this inertness, can play a significant causal role, in virtue of being instantiated in particulars which cause things. However, even if we grant that universals can play such a role in causation, this concession will not help Property-ist Platonism. The reason is that universals can have causal relevance in this way only through particulars in which they are instantiated. For example, the universal mass could be said to play a role in an event where a particular causes something in virtue of its mass, whereas the universal mass does not play any causal role in a situation where no mass is involved. But when a subject has a hallucination, by definition, no perceived object in the causal chain leading to the subject's experience instantiates the experienced property. If a subject is hallucinating a red tomato, she does not perceive any red object. Thus, the universal redness cannot causally affect the subject in virtue of any causally effective particular. We may conclude that on Property-ist Platonism, there is no quality causation in hallucinations, that is, hallucinated phenomenal qualities cannot be causally efficacious. Next, I will turn to the task of showing that this is a problem for Property-ist Platonism, by giving two reasons.

The first reason is that it seems experienced phenomenal qualities can have causal effects. It seems they can cause judgments, thought and talk about them in subjects that experience them. We say "ouch" because of the pain qualities we feel. We think, make judgments and philosophize about phenomenal qualities, *because* of the qualities we have experienced.

But if phenomenal qualities can have effects when experienced, it follows they should also have effects when experienced in hallucinations. This has been emphasized by Brad Thompson:

There is a further problem of causal efficacy on this view, which can again be illustrated by considering the experience of pain. Unless one is willing to adopt epiphenomenalism, one wants to say that pains can cause various behaviors, such as exclamations of "ouch" or the seeking of medication. In such cases, it is the phenomenal property associated with pain - the "painfulness" of the pain, that plays some role in effecting behavior. Uninstantiated properties simply cannot play that causal role. ...

[L]ying behind the intuition that pains and other conscious experiences are causally efficacious is the idea that such qualities can literally cause bodily motion. The kind of epiphenomenalism about consciousness that most philosophers seek to avoid is the view that denies this. Uninstantiated properties would seem to be, because they are not instantiated, incapable of this sort of causal efficacy (Thompson, 2008, 404).<sup>135</sup>

Here Thompson discusses the causal effects of experiences, focusing on the example of pain. These effects are caused by the "feel" of these experiences, or in other words, their phenomenal properties, which in this case are phenomenal pain qualities. When we feel pains we take it that it is their pain qualities which cause our utterances and thoughts about the badness of the pains. We take it that they have effects on our behavior, causing bodily motions. Thompson also claims that the position that phenomenal qualities lack causal effects is a type of epiphenomenalism, which most philosophers try to avoid. According to William Robinson, the idea that things like pains have causal effects might even be one of the main sources of resistance to epiphenomenalism in general:

Epiphenomenalism is absurd; it is just plain obvious that our pains, our thoughts, and our feelings make a difference to our (evidently physical) behavior; it is impossible to believe that all our behavior could be just as it is even if there were no pains, thoughts, or feelings....

This argument is surely the briefest of those against epiphenomenalism, but it may have been more persuasive than any other (Robinson 2012).<sup>136</sup>

We can see that lack of quality causation in hallucinations is problematic more clearly, if we look at a subject who has never been in contact with, or experienced, any phenomenal redness. To make the case clear, consider a

---

<sup>135</sup>Lycan (2001) also briefly discusses the intuition that phenomenal qualities should have causal effects on our behavior.

<sup>136</sup>Note that Robinson is not endorsing this argument but just describing it. According to Chalmers (1996, 159) this is the most common objection to epiphenomenalism. Unlike Robinson, Chalmers primarily discusses this intuition as an objection against epiphenomenalism about phenomenal qualities, rather than as an objection against epiphenomenalism in general. Chalmers also agrees that this is a strong intuition.

world where nothing is, will be, or has been red. It is possible that a subject in this world hallucinates red. To illustrate: The world may contain a green object. If a subject stares at this green object for a while and then looks at a white surface, she will experience a red afterimage even though neither the surface nor anything else in this world is red.<sup>137</sup> When having this experience she is having a hallucination where she experiences phenomenal red qualities. Now, she might learn from this hallucination what phenomenal redness is like. After her hallucination, she makes judgments and statements about phenomenal redness, based on her hallucinatory experience.<sup>138</sup> For instance, she describes how it contrasts with other phenomenal colors, and claims that the color feels “warm”. It here seems that the phenomenally red qualities the subject experienced should be involved when we give a causal explanation of her judgments and thoughts about phenomenal redness that she made. We should note however that lack of quality causation in hallucinations does not amount to general epiphenomenalism, such as to the epiphenomenalism of Thomas Huxley, which claims that the whole mind is epiphenomenal, including thoughts, beliefs, and volitions.<sup>139</sup> Even if phenomenal qualities lack causal effects, mental events such as thoughts, beliefs and volitions can still have causal effects, even in virtue of their mental properties.<sup>140</sup> *Experiences themselves*, even hallucinations, can also have causal effects. It is just that the phenomenal qualities experienced in them cannot have any causal effects.

Another difference between this problem and more general epiphenomenalism is that for EN views, which claim that phenomenal qualities experienced in veridical perceptions are properties of external objects, this problem only concerns causal effects of qualities experienced in hallucinations, not effects of qualities experienced in veridical perceptions. It might seem to be less of a loss, and thus less problematic, if only the qualities we experience in hallucinations lack causal effects on us. On the other hand, for Property-ist Platonist NN views, which claim that experienced phenomenal qualities are uninstantiated properties both in veridical

---

<sup>137</sup> We can assume that other than this, this possible world is just like ours. The brains of these subjects in this world are capable of perceiving redness, just as our brains. It is just that the subjects in this world have never seen anything red, since no object is in fact red there.

<sup>138</sup> Note that the possibility of such a world is not incompatible with the instantiation thesis, as it might seem on first sight. In this world nothing is red, that is, the external property redness is not instantiated. According to the instantiation thesis, since phenomenal redness is experienced, the property phenomenal red must be instantiated. But the property redness is, by hypothesis, nowhere instantiated in this world. However, these two states of affairs are only incompatible if phenomenal redness is identical to the external property redness, something, for example, brain theory denies.

<sup>139</sup> See Huxley (1999).

<sup>140</sup> For example Kim (2006, 304) and Fodor (1990, 156) seem to believe that epiphenomenalism about cognitive states is a lot worse than epiphenomenalism about phenomenal character. Proponents of epiphenomenalism about phenomenal character include Chalmers (1996), Jackson (1982), Jackson (1986) and Kim (2005).

perceptions and hallucinations, there is neither quality causation in veridical perceptions nor in hallucinations, so the problem concerns both types of experiences.

An objection to lack of quality causation being problematic is that one might think it is unfair to accuse a position which claims that the phenomenal qualities experienced in hallucinations are uninstantiated for making these qualities epiphenomenal. If the qualities are not instantiated, is it really a problem that they have no effects? That there are non-existing entities that do not affect the actual world causally does not make epiphenomenalism true, in the sense that it would entail that there are actual epiphenomenal entities. For example, if there is a non-existing Santa Claus, which does not have any causal effects upon the actual world, this does not make epiphenomenalism true. Similarly, the fact that a certain property which is not instantiated anywhere in the actual world does not affect anything in the actual world does not make epiphenomenalism true. For example, that the property of being a Jupiter sized ball of uranium is not instantiated anywhere, while at the same time not having any causal effects, does not make epiphenomenalism true.

However, this objection is beside the point, for two reasons. First, we should note that on Property-ism the phenomenal qualities are not non-existing. On Platonist Property-ism in particular, they are existing abstract objects. Thus, the entities which lack causal effects in this case are existing entities. Second, as I noted above, the problem with lack of quality causation does not concern the question of whether epiphenomenalism is true, but rather whether the phenomenal qualities experienced in hallucinations cause our judgments and thoughts about these qualities. If these qualities do not cause our judgments and thoughts this is problematic, regardless of whether the qualities happen to have “uninstantiated” or “non-existing” as their metaphysical status. So it is not wrong to use the quality causation objection to object against positions which claim that the phenomenal qualities experienced in hallucinations are uninstantiated or non-existing, as long as these positions claim that the qualities are experienced.

A second reason why lack of quality causation is problematic is that it makes it harder to explain how we can gain knowledge of what phenomenal qualities are like by having hallucinations. A plausible way of explaining how we can get knowledge of what something is like involves us being causally related to it. But this model of explanation is ruled out if there is no quality causation in hallucinations. Furthermore, if a causal theory of knowledge is

correct<sup>141</sup> then it follows directly that lack of quality causation is problematic, given that we can learn what phenomenal qualities are like through a hallucination.<sup>142</sup>

### 7.3 Aristotelianism

According to *Aristotelianism* (immanent realism) properties are universals, and universals exist in and through their instances.<sup>143</sup> For example, the universal brownness exists in all brown things. Contrary to Platonism, there is no separate entity Brownness distinct from all the brown things. A consequence of this dependence of universals on their instances is that if there are no brown things at a certain moment in time, then there is no universal brownness at that time. Since properties are universals on Aristotelianism, it also follows that there is no property brownness. Contrary to the trope theory, which is described below, Aristotelians take universals to be *wholly located* in each or their instances. Thus, universals are entities that can be multiply located, a “one that goes through the many” metaphorically speaking. If two red apples are located five meters apart it follows on this position that redness is an entity that is located both at one location and also five meters away from itself.

We can define the combination of Aristotelianism and Property-ism as follows:

---

<sup>141</sup> Such theories need not be rare. As Robert Kirk (2012) expresses it: "According to the widely accepted causal theory of reference - accepted by many philosophers - reference and knowledge require us to be causally affected by what is known or referred to."

<sup>142</sup> One argument that instantiation of qualities is not necessary to explain knowledge of phenomenal qualities derives from the Swampman thought experiment. Assume that by a strange accident a normal, living human being is suddenly created in a black-and-white swamp. Assume furthermore that Swampman is physically identical to me. I know what it is like to experience red, and it might seem like Swampman should also know what it is like to experience red, since he is physically identical to me.

Personally, it seems to me that having experienced phenomenal qualities is constitutive for having knowledge about phenomenal qualities. But I confess that this might be a very controversial reply. If this reply cannot be made, does the swampman objection also undermine the more general quality causation objection? I believe not. Although it might be possible in exceptional cases, such as the swamp man case, to make judgments about phenomenal redness without having had any experience of phenomenal redness, it seems like phenomenal redness still plays a causal role in normal cases of subjects who make judgments about phenomenal redness on the basis hallucinations.

<sup>143</sup>Armstrong (1989) is a famous proponent of immanent realism.

*Property-ist-Aristotelianism:* Aristotelianism is true and phenomenal qualities experienced in hallucinations are uninstantiated properties.<sup>144</sup>

Before criticizing this view, let us look at what it implies for phenomenal qualities experienced in hallucinations. Recall, on Property-ism these qualities are uninstantiated properties. Aristotelianism has a harder time giving a straightforward account for what uninstantiated properties is than Platonism had. This is because on Aristotelianism, properties are wholly located in their instances, so that there are no properties that are independent from their instances. In the context of hallucinations we should understand “uninstantiated properties” as properties which are not instantiated in an object the subject of the hallucination perceives. Since properties exist in their instances, it follows that these uninstantiated properties must be instantiated in some object that is not perceived by the subject of the hallucination. However, Property-ist Aristotelianism does not specify which unperceived object or objects the phenomenal qualities we experience in hallucinations exist in. Thus, the phenomenal qualities we experience in a hallucination exist in one, some, or perhaps even all of the unperceived instances of that property.

For example, in a case where the subject of a hallucination experiences phenomenal redness, according to Aristotelianism, the property redness exists in, one, some or all unperceived red things. Thus, when I have a hallucination of redness, the redness quality I experience might either be the redness of one of these things, the redness of some of them or the redness of all red things in the universe.

We can sum up the specific location Property-ist Aristotelianism gives to phenomenal qualities experienced in hallucinations as follows:

Phenomenal qualities experienced in a hallucination belong to some external objects which are unperceived by the subject of the hallucination, and instantiate corresponding external candidates.

Here “some” is used in a wide sense, so that in the case of redness, for example, it could refer to anything from a single red object to all red objects.

---

<sup>144</sup> Aristotelianism denies that a property can be uninstantiated in the sense that it is no instances whatsoever. But note that Aristotelianism does not deny that a property can be uninstantiated *relative to a certain hallucination*, in the sense that the phenomenal qualities experienced in that hallucination are not instantiated in a perceived object or in an internal object. Thus, the Property-ism Aristotelian claim is not inconsistent.

I have added the “external candidates clause” to restrict the relevant unperceived objects to objects which have the properties the Property-ist wants to identify the phenomenal qualities with. Recall, the “external candidates” are the external properties that are the best candidates for being the sensible qualities. For example in the case of a hallucination of phenomenal redness the relevant objects are red objects, objects which instantiate certain surface properties, which are the best candidates among external properties for being the property redness.

Let us proceed to our evaluation of the combined view. Property-ist Aristotelianism might strike some as very counterintuitive. For example, Thompson (2008), when criticizing this position, seems to take this counterintuitiveness in itself as a reason to reject the position:

If universals are taken to be Aristotelian, then this response does not resolve the challenge. There is no instance of the universal redness before Mary when she has a hallucination of redness. And surely what Mary experiences is not all of the instantiations of redness spread out across the universe, or some arbitrary instance of redness that is not before her (Thompson, 2008, 400).

On this view, the redness that Mary experiences while hallucinating is not the property of representing physical redness. It is not simply, like the weight of an object, a relational property of her experience that depends on relations to the external environment. The redness that is phenomenally manifest to Mary is entirely external, even in the case of hallucination. The redness that she experiences is the sum total of instances of redness scattered across the universe, or some arbitrary instance of redness. This, I think, is phenomenal externalism gone mad (Ibid. 402).

Thompson here notes the alternatives open to the Property-ist Aristotelian with disapproval, presumably for their counterintuitiveness. I agree with him that Property-ist Aristotelianism is indeed a rather peculiar view. According to it, when having a simple hallucination we experience far away objects that we have no causal contact with, or even an infinite amount of objects spread out across the whole universe.<sup>145</sup> However, I will not take this counterintuitiveness to be a sufficient reason to reject the position. Instead, I will focus on two other problems of Property-ist Aristotelianism.

The first problem is that the Property-ist Aristotelian has *too many* different alternatives open when specifying which unperceived objects experienced

---

<sup>145</sup> Why infinite? Contemporary cosmological models and physical theories seem to imply that there is an infinite amount of physical objects. See Knobe, Olum and Vilenkin (2006) and Tegmark (2003) for arguments to this conclusion.

phenomenal qualities belong to, and no non-arbitrary way to choose between these alternatives. The problematic consequence of this is that what objects hallucinated phenomenal qualities belong to must be determined by brute facts. To see this, let us once again consider a subject hallucinating phenomenal redness. On Property-ist Aristotelianism, the experienced redness has to belong to some object or objects. What object? It could be all the red objects in the universe, but it could also be a single red object, or some other group of red objects. Possible object(s) include everything from a random red object, the nearest red object, the red object that is most similar to the object that is represented by the experience, the red object the subject last saw and so on. Not only has the Property-ist Aristotelian not given us any good criterion to decide which object(s) the quality belongs to, but even worse, it seems no non-arbitrary criterion can be given. The reason is that the Property-ist Aristotelian has not postulated any mechanism that puts us in touch with unperceived objects. This is perhaps expected, since presumably no one would want to be committed to there being mechanisms connecting hallucinating perceivers with, perhaps far-away, unperceived objects. But without such a mechanism there are no facts which can ground the fact about which object(s) the redness experienced in hallucinations belongs to.<sup>146</sup> Rather, which object or objects the experienced redness belongs to must depend on a brute, unknowable fact. Since the postulation of such brute unknowable facts is something we should avoid if possible, this is problematic for Property-ist Aristotelianism. Let us call this objection the *brute fact problem* for Property-ist Aristotelianism.

The second problem, which also is my main argument against Property-ist-Aristotelianism, is that there are some hallucinations which the position is not applicable to, namely hallucinations of types of phenomenal qualities that are not instantiated anywhere. For example, take a world in which a person has a hallucination of redness, but no object in the whole world is red. Such a world might be unlikely, but it is possible.<sup>147</sup> According to Aristotelianism, since no things are red, there is no property redness in this world. Since Property-ists do not deny that we experience phenomenal qualities in hallucinations, this hallucinating subject experiences the quality phenomenal redness. However, Property-ism cannot explain the phenomenal qualities experienced in this hallucination the same way it

---

<sup>146</sup> Note that we cannot just stipulate an arbitrary criterion. If we experience a quality that belongs to some object then a fact about which object it belongs to must obtain prior to our philosophizing about the issue, and thus prior to any act of stipulation.

<sup>147</sup> A world without anything red might be ruled out by our laws of nature, at least if it contains a hallucinating perceiver. For example, a living human perceiver has to have blood pumping, and, perhaps due to its chemical nature, blood must be red to fulfill its role in the body. However, merely possible worlds need not have the same laws of nature as our world, so nothing prevents that there are other possible worlds, with different laws of nature, where blood has another color.

explains qualities experienced in other hallucinations. The reason is that in other hallucinations, according to the Aristotelianism Property-ist, the hallucinated experienced qualities belong to other unperceived objects, but in this scenario there are no other red objects in the whole universe that the experienced redness can belong to. Thus, some other theory than Property-ist Aristotelianism will be needed to account for the phenomenal qualities experienced in such hallucinations.

This need for another theory gives rise to a dilemma for Property-ist Aristotelianism. If there is no such other theory, then Property-ist Aristotelianism is unacceptable, since it does not apply to all hallucinations, and it is an incomplete theory of phenomenal qualities in hallucinations. On the other hand, if there is such a theory it seems we could apply that theory to hallucinations across the board and not just to the special hallucinations that Property-ist Aristotelianism fails to apply to, implying that there is no need for Property-ist Aristotelianism in the first place. It seems that out of the two horns of the dilemma, the least objectionable is that of being inapplicable to some hallucinations. Thus, we can call this problem the *inapplicability problem* for Property-ist Aristotelianism.<sup>148</sup>

A Property-ist Aristotelian might object that she can account for hallucinations of a property that is not instantiated anywhere if an instance of that property, and thus the property itself, has existed in the past, or will exist in the future. On Aristotelianism, universals exist in their instances, so a universal cannot exist *now* if it has no instances now, regardless of whether it had instances at some earlier time or not. However, the Property-ist Aristotelianism could claim that, although they do not exist now, the qualities we experience in hallucinations are located in past or the future.<sup>149</sup>

However, this line of reply does not help against the inapplicability problem. The reason is that just as a universe where a certain property is not instantiated anywhere at a certain moment in time is possible, likewise, a universe where a certain property is *never instantiated anywhere* at any time is also possible.

---

<sup>148</sup> Thompson (2008) uses an argument based on this implication against the thesis that phenomenal properties are Aristotelian universals.

<sup>149</sup> This might seem like a bizarre proposal, but note that, in the context of discussions of the so-called time gap argument for indirect perception, it has sometimes been claimed that we can perceive objects that existed in the past, but do not exist anymore. For example, if we see the light of a star located many light years away, and the star exploded a year ago, they argue that we could still be said to see the star even if it does not exist. What we see is located in the past. Huemer (2011), and Langsam (1997), among others, discuss the time gap argument for indirect perception and this line of response.

## 7.4 Nominalism

According to nominalism, there are no universals.<sup>150</sup> On *eliminativist nominalism*, there are no properties but only the objects we think have the properties. Instead, particulars are supposed to play to roles that properties are usually taken to play. Thus, terms like ‘redness’ and ‘squareness’ refer to nothing and there is no such thing as properties in the ontology. For example, there is no redness but only red objects, and there is no squareness, but only square objects. These objects may count as being red or square either through the similarities they bear to each other or because the predicates “red” or “square” apply to them.<sup>151</sup> *Reductionist nominalism*, on the other hand, accepts that there are properties but identifies them with something other than universals, such as sums or classes of actual or possible particulars. Terms such as ‘redness’ then does not refer to a universal but instead to some particulars, such as all actual red things, all possible red things or the particular that is the class of all red things.<sup>152</sup>

We can define the combination of nominalism and Property-ism as follows:

*Property-ist-Nominalism:* Nominalism is true and phenomenal qualities experienced in hallucinations are uninstantiated properties.

Let us look at what Property-ist nominalism says about the phenomenal qualities experienced in hallucinations, by looking at the combination of Property-ism and eliminativist nominalism, as well as the combination of Property-ism and reductionist nominalism, in turn.

Since an eliminative nominalist does not include properties in her ontology and instead thinks that the roles properties are supposed to play can be played by particulars, she cannot use uninstantiated properties to account for phenomenal qualities experienced in hallucinations. Thus, eliminativist nominalism is incompatible with the Property-ist claim that phenomenal

---

<sup>150</sup> Nominalism about universals is a theory which is distinct from nominalism about abstract objects. The latter claims that there are no abstract objects whereas the former claims that there are no universals. Thus the former is compatible with the existence of abstract particulars, such as sets or numbers. On different versions of nominalism, see Rodriguez-Pereyra (2011).

<sup>151</sup> The former is often called “resemblance nominalism,” the latter “predicate nominalism.” On these versions of nominalism see Rodriguez-Pereyra (2011).

<sup>152</sup> For example, according to Lewis (1983) the term refers to the class of things which have the property, to a mereological nominalist the term instead refers to the sum or aggregate of the things which have that property. See Rodriguez-Pereyra (2011).

qualities experienced in hallucinations are uninstantiated properties. If eliminativist nominalism is true, Property-ism is false.

On reductionist nominalism properties are either aggregates (mereological sums) of objects or abstract objects (sets of objects). Thus, *uninstantiated properties* would also be aggregates of objects or abstract objects, on reductionist nominalism. In the former case, the phenomenal qualities experienced in hallucinations are *sums* of unperceived objects. More specifically, they are aggregates of objects which instantiate the experienced phenomenal qualities. Although, strictly speaking, the resulting view is distinct from Property-ist Aristotelianism, nevertheless it is similar enough for the same objections to apply to it. The reason for this is that both views take the phenomenal qualities to belong to the same aggregates of objects. On Property-ist Aristotelianism, the phenomenal qualities are properties existing in these objects, and on this version of reductionist Property-ist nominalism they are the objects themselves. Due to this similarity, both views have the same problems with brute facts and inapplicability. First, just as on Property-ist Aristotelianism, when a subject experiences a phenomenal quality in a hallucination there are too many different alternative sums of particulars that the experienced phenomenal quality can belong to, but no non-arbitrary way to choose among them. In fact, the alternatives are made up of precisely the same particulars as the quality would belong to if Property-ist Aristotelianism was true. Thus, reductionist Property-ist nominalism faces the brute fact problem.

Second, on reductive nominalism, in a world in which no object in the whole world is, has been, or will be, red there is no property redness, since the property redness is constituted by the sum of all red objects. Thus, reductionist Property-ist nominalism cannot account for the experience of phenomenal redness if a person has a hallucination of redness in such a world, in the same way that it accounts for experienced phenomenal qualities in other hallucinations. So, reductionist Property-ist nominalism faces the same inapplicability problem as Aristotelianism did.

On the second version of reductionist Property-ist nominalism, which claims that phenomenal qualities experienced in hallucinations are sets the same objections as I used against Property-ist Platonism also apply against Property-ist nominalism. Sets are not universals, but rather particulars, so reductionist nominalism might seem to be in a better position to account for the particularity of the phenomenology of hallucinations than Property-ist Platonism was. However, even if abstract sets are particulars, so that we experience particulars in hallucinations, these sets do not provide us with any *instantiations* of properties, and more importantly, in a situation where

a subject has a hallucination as of several instantiations of the same phenomenal quality, there is still only one set that the subject experiences. For example, when a subject hallucinates two tomatoes of the same red color, say red<sub>59</sub>, the subject still does not experience any instantiation of phenomenal red, much less two distinct instantiations, as the phenomenology of the experience presents. Thus, reductionist Property-ist nominalism fails to account for the particularity of hallucinations. Furthermore, since the quality causation objection against Property-ist Platonism was based on the fact that hallucinated qualities are abstract objects on that view, the quality causation objection also applies against the combination of reductive nominalism and Property-ism.

Interestingly, this view also has the same inapplicability and brute fact problems as Property-ist Aristotelianism, despite claiming that the phenomenal qualities experienced in hallucinations are abstract objects. The reason for this is on this version of reductionist nominalism phenomenal qualities are not *pure sets*, sets that are set theoretic constructions of nothing but the empty set, but rather sets of concrete objects. This dependence on concrete objects revives the inapplicability objection. For example, in a world where there are no red objects, neither are there any sets of red objects, and hallucinated phenomenal redness quality cannot be located in any set of red objects. So this position is inapplicable to some hallucinations. Also, since for each sum of objects there is a corresponding set of the same objects, when a subject hallucinates a phenomenal redness quality there will be at least as many different sets of red tropes that the quality can be located in, as there are sums of concrete objects, but no non-arbitrary way to choose among them.

## 7.5 Trope theory

The final theory of properties we shall look at is trope theory.<sup>153</sup> This is sometimes seen as a version of nominalism.<sup>154</sup> Its proponents deny the existence of irreducible universals and claim instead that “tropes” fulfill the roles of universals. A trope is an abstract particular, a property instance. Each thing which instantiates a certain property has a trope of that property. That trope is a “bit” or part of the property; metaphorically speaking tropes

---

<sup>153</sup> Proponents of trope theory include Bacon (1995) and Maurin (2002).

<sup>154</sup> See for example Rodriguez-Pereyra (2011) on trope theory as a version of nominalism.

are seen as the building blocks which compose properties.<sup>155</sup> For example, each red thing has some redness in it, a certain bit of the property redness. The redness of a particular red thing is a red trope.

The trope theorist can say different things about properties. On a first interpretation, a property is a resemblance class of tropes.<sup>156</sup> This statement can in turn be interpreted in two ways. First, the properties might be the *sums* of the tropes in such classes. For example, redness itself would be the sum of all red tropes. Second, properties might be *sets* of the tropes in such resemblance classes. Then redness would be the set of all red tropes. This yields two slightly different theories about properties. On a second interpretation, which gives rise to a third theory of properties, trope theory says that individual instantiations of properties are identical to individual tropes, but that there are no properties over and above these instantiations. In that case the redness of one apple would be numerically identical to a certain redness trope and the redness of another qualitatively identical apple would be numerically identical to another redness trope, but there would be no entity redness over and above these individual tropes, not even a reducible entity.

We can define the combination of trope theory and Property-ism as follows:

*Property-ist-tropism:* Trope theory is true and phenomenal qualities experienced in hallucinations are uninstantiated properties.

The two different theories of properties that are possible on trope theory, which were mentioned above, give rise to two different corresponding versions of Property-ist tropism. The first, which says that phenomenal qualities experienced in hallucinations are *sums* of tropes, is very similar to Property-ist Aristotelianism. Qualities experienced in hallucinations are located in precisely the same particulars, with the only difference being that in the former case they are properties existing wholly in the particulars whereas in the latter case they are tropes existing in the same particulars. For example, if a subject hallucinates a red phenomenal quality, on Property-ist Aristotelianism the quality would belong to all, or some, of the red objects. On this first version of Property-ist tropism it would instead belong to tropes in all, or some, of the red objects. Because of this similarity the same objections as applied to Property-ist Aristotelianism, the inapplicability and brute fact problems, also apply to this version of Property-ist tropism.

---

<sup>155</sup> See Bacon (2011) on tropes as “bits” of properties. According to some trope theorists not only properties but also objects consist of tropes. An individual is then a bundle of tropes whereas a property is a similarity class of tropes.

<sup>156</sup> For example according to Lowe (2002, 361) property terms refer to resemblance classes of tropes.

The second version of Property-ist tropism says that phenomenal qualities experienced in hallucinations are *sets* of tropes. This position faces the particularity, inapplicability, and brute fact problems, for the same reasons as the set version of reductionist Property-ist nominalism discussed above. The only differences between these views is that in the former case hallucinated phenomenal qualities are sets of particulars, whereas in the this latter case hallucinated phenomenal qualities are sets of tropes, located in the same particulars. Furthermore, since sets of tropes are abstract objects, and abstract objects are causally inert, the mental causation objection also applies to this version of Property-ist tropism.

## 7.6 Nothingism and Meinongianism

Non-existence-ism says that we experience phenomenal qualities in hallucinations but that these qualities do not exist. As I mentioned in Chapter 4, we can distinguish two types of Non-existence-ism: Nothingism and Meinongianism. Below, I will consider them in turn and point out several problems for both of these positions.

Nothing-ism says that phenomenal qualities experienced in hallucinations do not exist, *simpliciter*. First, note the particularity objection applies to this position, since on Nothingism there are no particulars that we experience, which could explain the particularity of hallucinatory experience. If there were non-existing particulars that we experienced in hallucinations, these would be Meinongian objects. However, since Nothingism, by assumption, denies that experienced qualities are Meinongian objects, it must also deny that we experience non-existing particulars. Second, the quality causation objection also applies to Nothingism. The reason for this is that the phenomenal qualities are non-existing entities on Nothingism, and non-existing entities cannot causally affect existing entities.<sup>157</sup>

Let us now turn to Meinongianism. It says that phenomenal qualities are experienced in hallucinations, and that these qualities are non-existing entities which "have being."

---

<sup>157</sup> One could argue that absences causing absences are an exception. For example, that a rock-climber does *not* die while climbing might be thought to be causally related to her rope *not* breaking. However, in such cases the absences cause other absences. What is at issue here is the causation of positive mental events, such as beliefs and judgments, and not causation of absences. Furthermore, one might also argue, as Armstrong (1999) does, that omissions are not real causes.

Unlike Nothingism, Meinongianism can explain the particularity of hallucinations. When a subject has hallucinations she experiences instantiations, namely Meinongian instantiations which have being but do not exist. However, it has the same problems with quality causation as Nothingism had. The reason is that the qualities are non-existing entities, just as on the combination of anti-realism and Nothingism above. As non-existing entities cannot have causal effects on existing entities, the phenomenal qualities cannot have any causal effects, on Meinongianism.<sup>158</sup>

Furthermore, we should note that Meinongianism about phenomenal qualities experienced in hallucinations naturally is committed to Meinongianism *about non-existing objects*, the view that there are objects which do not exist. That is a controversial view, to say the least.<sup>159</sup> In addition, Meinongianism might violate Occam's razor, the principle that we should not multiply entities, all else equal. If my replies to the objections to brain theory in Chapter 10 below are correct then we do not need to postulate Meinongian objects to account for phenomenal qualities experienced in hallucinations. So if there are not other pressing concerns that necessitate the postulation of Meinongian objects the position violates Occam's razor.

Finally, I will note, using the value-supervenience argument below, that both Meinongianism and Nothingism have a hard time explaining how phenomenal qualities experienced in hallucinations can have quality-dependent value depending on them

The value supervenience argument basically goes as follows. When a subject has a valuable hallucination, the quality-dependent value of the hallucination does not only depend on the phenomenal qualities experienced in it, but the value also *supervenies* on them. A set of properties P supervenes upon another set Q, just in the cases where no two things can differ with respect to P-properties without also differing with respect to their Q-properties. Since actual properties only supervene on existing things, and

---

<sup>158</sup> Also note that both Nothingism and Meinongianism seem to give rise to something similar to the brute fact problem. Take a subject which has a hallucination as of a red apple. Presumably there are very many different non-existing apples which look just like the hallucinated apple. The problem is that the qualities must belong to one of these non-existing apples, but not only do we have no way of knowing which apple the qualities belong to, it seems there are no relevant facts which can make it the case that the qualities belong to one non-existing apple rather than another. Thus, the view must postulate brute unknowable facts about precisely what non-existing objects hallucinated phenomenal qualities are identical to.

<sup>159</sup> According to Reicher (2010) some believe that the very concept of a non-existing object is contradictory or logically ill-formed, see Hume (2000, Book 1, Part 2, Sect. 6); Kant (2003, B 627) and Frege (1966, 37f). In addition there are those like Quine (1953) who deny that one can make a distinction between "There are Fs" and "Fs" exist.

this value is actual we reach the conclusion that these qualities must be actual, and thus that they exist.

More formally, we might now express *the value supervenience argument* as follows:

P1 Phenomenal qualities are experienced in some hallucinations that have quality-dependent value that depends on these qualities.

P2 If value supervenes on a quality, then that quality exists.

P3 Quality-dependent value of hallucinations supervenes on the qualities it depends on.

C1 (by P1 and P3) In some hallucinations, quality-dependent value supervenes on phenomenal qualities experienced in them.

C (By C1 and P2) Existing phenomenal qualities are experienced in some hallucinations.

The conclusion, C, follows from the premises, P1-P3. According to P1, phenomenal qualities experienced in some hallucinations have quality-dependent value depending on them. By P3 such value supervenes on the qualities it depends on. Thus, in some hallucinations quality-dependent value supervenes on phenomenal qualities experienced in them, as C1 claims. C1 mentions phenomenal qualities that have value supervening on them, by P2 it trivially follows that these phenomenal qualities exist. Since these qualities, according to C1, are experienced in hallucinations it follows that existing phenomenal qualities are experienced in some hallucinations and thus that the conclusion of the argument, C, is true.

Let us look at the premises in turn. P1 is a result from the last chapter, justified by the value argument and the knowledge-acquisition argument. The second premise, P2, says that if the quality-dependent value supervenes on a quality, then that quality exists. It seems hard to deny that if actual value supervenes on something, that on which it supervenes must exist. Actual causation requires the existence of the relata, and supervenience is an at least as intimate relation, so we should expect that an actual property could only supervene on something else that is actual.

The third premise, P3, says that quality-dependent value of hallucinations supervenes on the qualities on which it depends. This is the most controversial and important premise in the value supervenience argument.

After presenting the evidence for this premise I will present two objections to it, together with replies to the objections.

One way to support P3 is through the following comparison test. Take a valuable experience, for example a pleasant aesthetic experience of a magnificent forest landscape. Here various qualities are experienced: phenomenal colors and shapes, as well as pleasure qualities. The value of this experience can be changed solely by changing the phenomenal qualities that are experienced. For instance, if all the color qualities were replaced with pitch black and the pleasure qualities were replaced with pain qualities the experience would no longer have the same value. This fact gives us evidence that the quality-dependent value of experiences supervenes on phenomenal qualities, or at least that they are included in the supervenience base for the value.

A representationalist Non-existence-ist might object that the alternative representationalist hypothesis that the value depends on what the experience represents would predict the same thing, since a change in the qualities would amount to a change in what is represented, so that the hypothesis that the value supervenes on the phenomenal qualities in the experience is not favored over the hypothesis that the value instead supervenes on the *representational content* of the experience. However, note that on typical of representationalist EN and NN positions, the qualities experienced in an experience are representational content or representational properties included in the representational content. Thus, on such views, even if we concede the hypothesis that the value supervenes on the representational content, we still get the conclusion that it supervenes on the phenomenal qualities.

It can be argued that not only phenomenal qualities, but also mental properties of a subject, such as attitudes, beliefs and desires, can change the quality-dependent value of an experience. However, even if that is the case it just provides a reason for also including these mental properties in the supervenience base for quality-dependent value, not a reason for excluding the phenomenal qualities from the supervenience base of quality-dependent value.

A second way to support P3 is through introspection. When you introspect a pain experience, for instance, you can see that what you dislike and what is bad about the experience are the annoying pain qualities. We can see this more clearly by considering the so-called transparency phenomenon. According to it, when you try to introspect an experience, you only find

certain qualities; you cannot directly introspect experiences themselves or qualities of experiences.

Suppose you are facing a white wall, on which you see a bright red, round patch of paint. Suppose you are attending closely to the color and shape of the patch as well as the background. Now turn your attention from what you see out there in the world before you to your visual experience. Focus upon your awareness of the patch as opposed to the patch of which you are aware. Do you find yourself suddenly acquainted with new qualities, qualities that are intrinsic to your visual experience in the way that redness and roundness are qualities intrinsic to the patch of paint? According to some philosophers, the answer to this question is a resounding 'No'. As you look at the patch, you are aware of certain features out there in the world. When you turn your attention inwards to your experience of those features, you are aware that you are having an experience of a certain sort but you are aware of the very same features; no new features of your experience are revealed. In this way, your visual experience is transparent or diaphanous. When you try to examine it, you see right through it, as it were, to the qualities you were experiencing all along in being a subject of the experience, qualities your experience is of (Tye 2009).

Here, I take it that when Tye mentions colors and shapes he is talking about phenomenal qualities, which he takes to be properties of external objects. Interpreted as concerning phenomenal qualities the transparency observation seems correct. Regardless of how hard we try, we can merely introspect phenomenal qualities directly, not any "experiences" or qualities of experiences that are distinct from the phenomenal qualities. But now, given the plausible assumption that the value of experiences with quality-dependent value depends on something directly given, and that nothing other than phenomenal qualities are directly given in hallucinations, as seems to follow from the transparency observation, we reach the conclusion that this value supervenes on phenomenal qualities.

A first objection to P3 is that value supervenes on *the experiences themselves* but not on the qualities experienced in them. In a paper which inspired many of the modern transparency claims, G.E. Moore claimed that we can distinguish two elements in experience: "when we try to introspect the sensation of blue, all we can see is the blue: the other element is as if it were diaphanous" (Moore 1903, 450).<sup>160</sup> We might interpret this claim as saying that on the one hand there is the phenomenal qualities, and on the other hand there is a second element, which is in some sense transparent ("diaphanous"). We might aptly call the second, transparent, element "the

---

<sup>160</sup> For an example of a modern transparency claim that was influenced by Moore, see Tye (2000, 111-112).

experience itself". Thus, an alternative hypothesis would be that it is this second element, and not the phenomenal qualities, that the value of experiences supervenes on. However, by the transparency observation above it follows that properties of experiences, other than phenomenal qualities, if they belong to experiences, are not directly given. But quality-dependent value seems to depend on features that are directly given. Thus, it is not the case that quality-dependent value supervenes on the experiences themselves, as the alternative hypothesis claims, rather than the qualities experienced.

A second objection to P3 is that the value depends on the relation of *experiencing the qualities* rather than the experienced qualities themselves. This objection can be supported by pointing out that there is a link between quality-dependent value and subjects standing in experiencing relations to qualities. For example, pain qualities somehow existing on their own, without any subject to experience them, might not seem to be bad, whereas pain qualities that are experienced by a subject are bad.

However, even if the relation to the experiences is important to the value, the value also depends on the qualities, as we have seen above. The conclusion to draw from this objection, if it is correct, as in the case of the mental properties mentioned above, is not that the value does not supervene on phenomenal qualities, but rather that it supervenes on both the experienced phenomenal qualities and the fact that the subject stands in an experiencing relation to them, in other words, that the relation should be included in the supervenience base. Changes in the qualities can change the quality-dependent value, and so can a change in the relation to the qualities.

# 8. Physicalism

## 8.0 Introduction

After describing the formulation of physicalism that is most relevant for our purposes, type physicalism, I will look at the arguments for physicalism, focusing on the causal argument. Finally, I will briefly overview the most central objections to physicalism, and strike a blow for physicalism by giving replies to these objections.

## 8.1 Type physicalism

Physicalism is the thesis that everything is physical. There are different views on what this thesis means more specifically.<sup>161</sup> One such view is:

*Type physicalism (TP):* For every actually instantiated property F, there is some physical property G such that  $F=G$ .

When I use the term ‘physicalism’ I will use it to mean what is expressed by type physicalism. Note that type physicalism does not imply the truth of the Psycho-Physical brain theory, which says that each type of mental state is identical with a type of *neural* state. Type physicalism is compatible with Gs being functional property or other non-neural properties, as long as these properties are physical.

## 8.2 The arguments for physicalism

The main motivations for physicalism are causal considerations.<sup>162</sup> One way to formulate these is through the causal argument for physicalism, as expressed by David Papineau (2002). The argument goes as follows.<sup>163</sup> We have good empirical reasons to believe that the physical domain is causally

---

<sup>161</sup> See Stoljar (2009) for a good overview of these views.

<sup>162</sup> See Stoljar (2009) for an overview of the motivations for physicalism.

<sup>163</sup> See Papineau (2002). The argument is also discussed by Stoljar (2009) and Kim (2006).

closed, that is, that any physical effect has sufficient physical causes. But then it follows that, if we want to avoid both epiphenomenalism and massive overdetermination, conscious causes must be identical to physical causes. Overdetermination occurs when an effect E has two distinct causes, and each of these two causes is sufficient to cause E on its own. For example, the death of a man who is simultaneously struck by lightning and shot by a bullet is arguably an example of an overdetermined effect.

We might express the causal argument more formally, and in terms of phenomenal qualities, as follows:

*The causal argument:*

P1 Each phenomenal quality  $Q_n$  is sufficient cause for some physical effect  $E_n$ .

P2 All physical effects,  $E_n$ , are sufficiently caused by physical properties,  $P_n$ .

P3 The physical effects of phenomenal qualities are not always overdetermined by distinct sufficient causes.

C1 (by P1 and P2) Either  $Q_n=P_n$  or  $E_n$  is always overdetermined by the distinct sufficient causes  $Q_n$  and  $P_n$ .

C2 (by C1, P3)  $Q_n=P_n$

C3 (by C2, definition of type physicalism) Type physicalism is true about phenomenal qualities.

The conclusion follows from the premises, so let us look at the premises in turn to see whether they are true. The first premise, P1, says that phenomenal qualities are sufficient causes for physical effects. For example, experiencing a nasty phenomenal pain quality might lead to me saying "ouch" and taking a step away from the cause of the pain. The premise is equivalent to a denial of epiphenomenalism and parallelism about phenomenal qualities, and the plausibility of the premise rests on the implausibility of these positions.

Parallelism or "pre-established harmony" is a position on mental causation, which was championed by Leibniz. According to this position, there is no causal interaction between mental and physical events at all, not even in the physical-to-mental direction. There only seems to be such interactions

because God has arranged for the mental and physical realms to run in parallel. I will focus on epiphenomenalism, since parallelism rests on dubious theological assumptions, and has not been defended since the time of Leibniz, as far as I know.

Note that, just like the limited form of epiphenomenalism discussed in chapter 7, this epiphenomenalism is not full “Huxley style” epiphenomenalism, which claims that all mental states are epiphenomenal. Insofar as the causation of beliefs, desires and other mental states does not depend on phenomenal qualities having causal effects, these other mental states are not rendered epiphenomenal even if P1 is false. We might aptly call this limited form of epiphenomenalism *quality-epiphenomenalism*.

Why is quality epiphenomenalism implausible? First, there are the intuitions against epiphenomenalism that were mentioned above, in Chapter 7. According to these intuitions, experienced phenomenal qualities have causal effects. They cause judgments, thought and talk about themselves in subjects that experience them. Note here that quality epiphenomenalism is more general than the form of epiphenomenalism that was discussed in Chapter 7, since the former concerns phenomenal qualities experienced in both veridical perceptions and hallucinations, in contrast to the latter less general form of epiphenomenalism, which only concerns phenomenal qualities experienced in hallucinations.

Second, there is an additional argument against quality-epiphenomenalism, based on scientific methodology considerations. Papineau (2002) expresses this argument as follows, when arguing against epiphenomenalism about mental and conscious states:

If epiphenomenalism were true, then the relation between mind and brain would be like nothing else in nature. After all, science recognizes no other examples of ‘causal danglers’, ontologically independent states with causes but no effects. So, given the choice between epiphenomenalism and materialism, standard principles of scientific theory choice would seem to favor materialism. If both views can accommodate the empirical data equally well, then ordinary scientific methodology will advise us to adopt the simple view that unifies mind and brain, rather than the ontologically more profligate story which has the conscious states dangling impotently from the brain states (Papineau 2002, 23).

Applied to experienced phenomenal qualities the idea is that if quality epiphenomenalism is true, so that phenomenal qualities do not have any physical effects, these qualities would be unlike everything else that we have discovered in nature. But by ordinary scientific methodology considerations

we should then prefer a physicalist theory of phenomenal qualities to epiphenomenalism, as long as epiphenomenalism does not have any other advantages over that physicalist theory, since the physicalist theory allows phenomenal qualities to have physical effects. The second premise, P2, says that all physical events have a sufficient physical cause.<sup>164</sup> This thesis is sometimes called “causal closure”. For example, if I stretch out my arm to grab a glass of water the motion was caused by muscular contractions, which in turn were caused by nerve signals, which in turn were caused by physical events in the motor cortex of the brain, and so forth.

P2 is not supposed to be justified *a priori*, but rather has to be supported by empirical evidence. However, the premise seems to have a decent grounding in empirical findings. As science has progressed, we have discovered sufficient physical causes for more and more types of events that were previously thought to lack physical causes.<sup>165</sup> The final area of contention seems to be agency and consciousness, but our behavior is caused by our brains, and there seems to be no strong reasons to believe that the brain is not a fully physical system where each event has a sufficient physical cause. Rather, the more we have learned about the brain the more underlying physical causal mechanisms we have uncovered. Conversely, when investigating the brain we have found no place where physical effects happen but no physical causes can be found, as we should expect to be able to find if interactionist dualism was true.

Even if P1 and P2 are true, so that experienced phenomenal qualities sufficiently cause physical effects, and these effects have sufficient physical causes, all versions of phenomenal quality dualism are still not ruled out. The reason is that these effects could be systematically overdetermined, so that they were always independently sufficiently caused both by a non-physical experienced phenomenal quality *and* by a distinct previous physical cause. The third premise, P3, rules out the possibility that the effects of phenomenal qualities are always overdetermined, so together with P1 and P2 it entails that phenomenal qualities must be physical.

Papineau (2002) gives an argument against systematic overdetermination which is analogous to the previously mentioned argument against epiphenomenalism that was based on scientific methodology considerations. Just as we do not find any “causal danglers” in nature, we neither find any

---

<sup>164</sup> Note that this premise by itself does not rule out the possibility of physical events having non-physical causes, since this premise is compatible with physical events being overdetermined. However, the combination of P2 and P3, the premise which rules out systematic overdetermination, does rule out the possibility of physical events having non-physical causes.

<sup>165</sup> For a more thorough defense of physical causal closure, see Papineau (2002).

cases of mechanisms that ensure that certain effects are always overdetermined. Thus, by standard principles of theory choice, we should choose the physicalist model of mental causation over that of the proponent of overdetermination, as long as they both have the same explanatory capacities.

A further reason for P3 is, as Kim (2006, 196) points out, that if we allow that the effects of experienced phenomenal qualities are systematically overdetermined their status as genuine causes is weakened. The reason for this is that, given P3 these effects have sufficient physical causes, and would thus happen even if phenomenal qualities were never experienced.

### **8.3 The anti-physicalist arguments**

Physicalism faces some considerable challenges. I will here focus on two of the most discussed challenges: the “knowledge argument” and the “conceivability argument” (also known as “the zombie argument”). Both of them have been met with various, often parallel, replies from physicalists. Below, I will briefly overview the anti-physicalism arguments and strike a blow for physicalism, by giving some physicalist replies.

The knowledge argument asks us to conceive of Mary, an exceptionally talented color scientist who has been raised in a completely black and white environment and thus never seen any color at all. In fact, after her education is complete (attained through a black and white TV set) she knows everything physical there is to know about the human visual system and the physical of color. She is now let free and sees color for the first time in her life; perhaps a bright red rose. As the argument goes, upon seeing the red rose she will *learn* something new: what it is like to see red. Since she already had all physical knowledge, this new knowledge cannot be about physical facts. But then there must be facts which are not physical, and thus physicalism is false.

Many replies to the knowledge argument have been proposed.<sup>166</sup> The simplest type of reply is to claim that Mary can know what red is like even prior to her release. For example, Dennett (1991) argues that the intuition that Mary learns something when she first sees red is wrong. When we

---

<sup>166</sup> For an overview that emphasizes this fact see Van Gulick (2005). For other overviews of the debates around the knowledge argument see also Nida-Rümelin (2010).

conduct the Mary thought experiment, we underestimate what it is to know *all* physical truths about the human visual system. According to Dennett, a person with such complete knowledge would be able to deduce truths such as what it is like to see red, even if she, like Mary, had never in fact seen red. If she already knows what red is like when being released, she does not learn anything new, and we have no reason to believe that there are non-physical facts.<sup>167</sup>

A second type of reply argues that Mary learns something when set free, but that what she learns is a new ability, such as the ability to imagine redness, rather than a new fact. If Mary does not learn any new fact when experiencing red the conclusion that there are non-physical facts does not follow.<sup>168</sup>

Others have argued that what Mary learns when she experiences redness is merely a new concept, a phenomenal concept, which lets her represent facts about seeing red that she already knew, but in a new way. Thus, there is no new fact that she learns when she is released, and the Mary thought experiment fails to justify the conclusion that there are non-physical facts.<sup>169</sup>

Finally, another type of reply claims that there are physical facts which cannot be learned without having certain experiences, and that what it is like to experience red is one such fact. Then, what she learns can be a physical fact after all. I myself favor the last line of response.

One version of this response, developed by Stoljar (2001a, 2001b), focuses on the distinction between dispositional physical properties and their categorical bases.<sup>170</sup> This reply rests on two assumptions. The first is that dispositional physical properties must have categorical bases.<sup>171</sup> For example, in classical physics particles with a certain mass are disposed to behave in certain ways. Reasonably, there is a categorical property these particles have, *mass*, which grounds this disposition. The second assumption is that physical sciences only give us knowledge about the dispositional properties, but not their categorical bases. For example, physics does not tell us anything about what *mass* is, except that it makes particles behave in certain ways. These tendencies to behave in certain ways are the dispositional properties that the categorical property mass grounds. Thus, physics only

---

<sup>167</sup> Dennett (1991) argues for this solution.

<sup>168</sup> See Lewis (1988b) for a presentation of this so called "ability hypothesis."

<sup>169</sup> For example, Papineau (2002).

<sup>170</sup> Stoljar (2006) defends and develops a more general type of the reply that this reply is a version of: that there are physical truths relevant to the nature of experience that we are ignorant of.

<sup>171</sup> The thesis that dispositional properties must have categorical bases is defended by Armstrong (1968) and Blackburn (1992), among others.

provides knowledge about the dispositional properties that mass grounds, not about mass itself. By the second assumption it follows that Mary can only learn about physical dispositional properties while confined to her room. But by the second assumption it follows that there are also other physical properties, the categorical bases, which Mary is ignorant about, despite having received a complete education about physical theory through a black and white television set. Thus, if truths about what it is like to experience red concern categorical physical properties, rather than dispositional physical properties, Mary can fail to know what it is like to experience red prior to experiencing red, and it does not follow that physicalism is false even if Mary learns something when she first experiences red.

We might call this line of response “Russellian”, since Russell (1927, 384) made an early statement of the idea that the categorical properties of physical objects are unknown, and that this ignorance would remove objections to the thesis that experiences, or "percepts" as Russell called them, are physical:<sup>172</sup>

A piece of matter is a logical structure composed of events: the causal laws of the events concerned, and the abstract logical properties of their spatio-temporal relations, are more or less known, but their intrinsic character is not known. Percepts fit into the same causal scheme as physical events, and are not known to have any intrinsic character which physical events cannot have, since we do not know of any intrinsic character which could be incompatible with the logical properties that physics assigns to physical events. There is therefore no ground for supposing that percepts cannot be physical events, or for supposing that they are never compresent with other events. (Russell 1927, 384)

The conceivability argument asks us to imagine a philosophical zombie – a creature physically qualitatively identical with the living body of a human being but completely without any phenomenal character. The argument goes as follows, in one, relatively simple, formulation. Since what is conceivable is possible, it follows that zombies are possible. Thus, it is possible phenomenal character is distinct from physical properties. But since identity is a necessary relation, states which are identical must be identical in all possible worlds. So, we arrive at the conclusion that phenomenal character is *actually* distinct from physical properties, which implies that physicalism is false.<sup>173</sup>

---

<sup>172</sup> This connection has been pointed out by Stoljar (2006, 107).

<sup>173</sup> For a classical presentation of the zombie argument see Chalmers (1996). An updated version of the argument is given in Chalmers 2009. See also Kripke (1972/1980).

A central line of reply to the conceivability argument focuses on denying that conceivability entails possibility.<sup>174</sup> Furthermore, some have denied that zombies are conceivable.<sup>175</sup> Just like in the case of the knowledge argument, we may note that brain theory does not preclude either of these physicalist replies to the argument.

The reply to the conceivability argument that I myself favor is a Russellian reply, which parallels the previously mentioned Russellian reply to the knowledge argument. Briefly, this reply says that when we seem to conceive of a zombie, which is not itself phenomenally conscious, but physically identical to a phenomenally conscious person, we only conceive of a creature which has the same dispositional physical properties as the person, not a creature which has the same categorical physical properties. The reason we cannot conceive of a creature with precisely the same categorical properties is that it follows that we are ignorant of these categorical properties from the assumption that physical theory does not provide knowledge of categorical bases.<sup>176</sup> Thus, the conceivability argument can only show that phenomenal qualities are distinct from dispositional physical properties, not that they are distinct from categorical physical properties.

---

<sup>174</sup> See e.g. Hill and McLaughlin 1999, Hill 1997 and Loar 1990/1997.

<sup>175</sup> E.g. Dennett (1995).

<sup>176</sup> If phenomenal properties are such categorical physical properties we are not ignorant of them, since we can learn about them from experience. However, in that case we cannot conceive of a creature which both has these properties and lacks phenomenal properties, since we cannot conceive of a creature which both has and lacks phenomenal properties.

## 9. Generalizing to the veridical case

### 9.0 Introduction

In this chapter I argue for the conditional claim that brain theory (BT) can be reasonably inferred from the instantiation thesis (IT), the hallucination\* thesis (H\*T), the experience thesis (ET) and type physicalism, taken together. Recall, these theses state the following:

*Brain theory (BT):* For all subjects S and phenomenal qualities Q: If S experiences Q, there is a physical property P, such that P is instantiated in the brain of S and  $P=Q$ .

*Instantiation thesis (IT):* For all subjects S and phenomenal qualities Q: If S experiences Q in a hallucination, then there is some existing internal object x, such that x is Q.

*Experience thesis (ET):* We experience phenomenal qualities in hallucinations.

*Hallucination\* thesis (H\*T):* For each veridical perception P there is a corresponding indistinguishable possible hallucination\* H\*

*Type physicalism (TP):* For every actually instantiated property F, there is some physical property G such that  $F=G$ .

I do this in three steps. First, I show how it follows from IT and type physicalism that if phenomenal qualities are experienced in hallucinations these qualities are instantiated in the brain of the subject of the experience. Second, I use the “identity argument” to generalize this conclusion about hallucinations to veridical perceptions. This gives us the conclusion that for every phenomenal quality you experience in a veridical perception, there is an indistinguishable quality instantiated in your brain. The third step argues that the experienced quality is identical with the quality in your brain, and so, together with the results about hallucinations, gives us brain theory as a conclusion.

In addition, I will briefly sketch a version of the classic causal generalizing argument, originating from Robinson (1985) and compare it with the identity argument. The reason I also include this argument sketch is to illustrate that even if physicalism is false, it might still be possible to

generalize the results of ET, IT, and H\*T about hallucinations to veridical perceptions. However, since the causal generalizing argument is not part my main argumentation for brain theory, I will not fully investigate the truth of the premises of this causal generalizing argument, in particular the “same sufficient proximal cause, same effect” premise. Because of this we cannot use it to support brain theory.

### **9.1 The first step**

Let us begin with the first step. According to IT phenomenal qualities experienced in hallucinations must be instantiated in existing internal objects. The internal objects are either located in the brain or in non-physical sense data. However, by TP there are no non-physical objects. Thus, if hallucinating subjects experience phenomenal qualities these qualities are instantiated in their brains. We can conclude that it follows from IT and TP that when a subject experiences hallucinated phenomenal qualities, these qualities are instantiated in her brain:

*Hallucination brain thesis (HBT)* For all subjects S and phenomenal qualities Q: If S experiences Q in a hallucination, then Q is instantiated in the brain of S.

Note also that because of the TP assumption these phenomenal qualities that are instantiated in the brains of subjects who experience hallucinations must be physical properties.

The only alternative to brain theory that is compatible with the conclusion HBT is EI brain theory:

*EI brain theory:* All phenomenal qualities experienced in hallucinations are instantiated in the brain. Phenomenal qualities experienced in veridical perceptions are not instantiated in the brain.

Since EI brain theory denies that phenomenal qualities experienced in veridical perceptions are instantiated in the brain it is incompatible with brain theory, and will be ruled out by the conclusion of this chapter. Note that the identity argument below is not enough on its own to entail that this position is false, since it is possible to agree that phenomenal qualities are instantiated in our brains when we have veridical perceptions, while at the same time denying that these qualities are experienced. However, the third step is sufficient to rule out EI brain theory.

## 9.2 The identity argument

Briefly, the identity argument goes as follows. Take an arbitrary phenomenal quality  $Q$  experienced in a veridical perception  $V$  by the subject  $S$ . By the hallucination\* thesis, for each veridical perception there is a possible corresponding hallucination\*  $H^*$ . By the experience thesis,  $S$  experiences phenomenal qualities in  $H^*$ . As  $H^*$  is indistinguishable from  $V^*$ , among these qualities is a quality,  $Q^*$ , which is indistinguishable from  $Q$ . Since, as we saw above, phenomenal qualities experienced in hallucinations are instantiated in the brain,  $Q^*$  is instantiated in the brain of  $S$ . By type physicalism, it follows that these qualities are identical to physical properties. So  $Q^*$  is identical to a physical property,  $G$ . Since  $Q^*$  is instantiated in the brain of  $S$  it follows by the indiscernibility of identicals, the principle that identical entities cannot differ in any respect, that  $G$  must also be instantiated in the brain of  $S$ . Now, by the definition of the term 'hallucination\*',  $S$  is in the same global physical brain state when having  $V$  as when having the  $V$ 's corresponding indistinguishable hallucination\*  $H^*$ . Among other properties,  $S$ 's brain thus instantiates  $G$  when  $S$  has  $V$ . But now recall that  $G$  is identical to  $Q^*$ . This gives the conclusion that  $Q^*$  is instantiated in the brain of  $S$  also in  $V$ . As  $Q$  was an arbitrary phenomenal quality experienced in an arbitrary veridical perception we can draw the general conclusion that for every phenomenal quality you experience in a veridical perception, there is an indistinguishable quality instantiated in your brain, and that quality is a physical property of your brain.

More formally:

P1 (H\*T) For each veridical perception  $P$  there is a corresponding indistinguishable possible hallucination\*,  $H^*$

P2 (ET) We experience phenomenal qualities in hallucinations.

P3 (HBT) For all subjects  $S$  and phenomenal qualities  $Q$ : If  $S$  experiences  $Q$  in a hallucination, then  $Q$  is instantiated in the brain of  $S$ .

P4 (TP) For every actually instantiated property,  $F$ , there is some physical property  $G$  such that  $F=G$ .

C1 (by H\*T) For an arbitrary veridical perception, V, had by the subject S, where the arbitrary phenomenal quality Q is experienced, there is a corresponding possible hallucination\*, H\*.

C2 (by, C1 and ET) A phenomenal quality Q\* is experienced in H\*, and Q\* is indistinguishable from Q.

C3 (by C2 and HBT) The phenomenal quality Q\* is instantiated in the brain of S in H\*.

C4 (by TP and C3) There is a physical property G, such that  $G=Q^*$ .

C5 (by C3, C4 and the indiscernibility of identicals) G is instantiated in the brain of S in H\*.

C6 (by C5 and the definition of 'hallucination\*') G is instantiated in the brain of S in V.

C7 (by C6 and C4) Q\* is instantiated in the brain of S in V and Q\* is identical to the physical property G.

C8 (by C1-C7) When a subject experiences a phenomenal quality Q in a veridical perception, there is an indistinguishable phenomenal quality Q\* instantiated in her brain, and Q\* is a physical property.

### 9.3 The third step

Let us proceed to the third step. It follows from HBT and TP that all phenomenal qualities experienced in hallucinations are physical properties instantiated in the brain. So brain theory follows from the four theses ET, IT, H\*T and TP if we can also show that *all* phenomenal qualities experienced in veridical perceptions are physical properties instantiated in the brain.

We know from the conclusion of the identity argument, C8, that when we experience a phenomenal quality Q in a veridical perception, an indistinguishable phenomenal quality Q\* is instantiated in our brain. The remaining question is whether Q and Q\* are identical. If they are, then brain theory follows.

I will give two reasons for the identity claim  $Q=Q^*$ . But let us first consider what follows if we deny this claim. In that case, whenever you have a

veridical perception there are two indistinguishable sets of qualities involved. One set is instantiated in your brain. But there is also second indistinguishable set of qualities, which you experience, that is not instantiated in your brain. Rather, these qualities have another location, plausibly instantiated in external objects. Let us call this the two sets view.

The first reason for that  $Q=Q^*$  is that economy considerations would seem to dictate that we should not postulate that there are two sets of phenomenal qualities involved in veridical perceptions, when one set is enough to explain the phenomenology of our veridical perceptions. Thus, when we experience a phenomenal quality  $Q$ , and know that an indistinguishable quality  $Q^*$  is instantiated in our brain, we should not postulate a second phenomenal quality  $E$ , which is distinct from  $Q^*$ , and has another location, such as being instantiated in an external object.

The second reason is that if we know that one set of phenomenal qualities are *strictly identical* to physical brain properties, for example neural properties, it is hard to see how an indistinguishable set of phenomenal qualities can be identical to perception-relevant properties of external objects, such as reflectance properties. For example, assume that the phenomenal quality phenomenal  $red_{59}$  is strictly identical to a type of activation pattern of brain neurons. Now, on the two sets view, there is a second phenomenal quality,  $red^*_{59}$ , that is indistinguishable from  $red_{59}$ , and external. A plausible external location for phenomenal red qualities, on views which take them to be external, is at the surfaces of objects like tomatoes, as reflectance properties. So let us assume that according to the two sets view  $red^*_{59}$  is a reflectance property of tomatoes. However, the physical reflectance properties of tomatoes are very dissimilar from physical patterns of activation among brain neurons. Thus, it is hard to see how two perfectly indistinguishable phenomenal qualities could be strictly identical to two so dissimilar things as reflectance properties and neural properties. If phenomenal qualities were non-physical, perhaps both tomatoes and brains could instantiate two different but indistinguishable phenomenal qualities. But remember that we have assumed type physicalism; all instantiated properties are strictly identical to physical properties.

To conclude this chapter, it seems the three steps provide a justification for brain theory on the basis of the four theses ET, IT, H\*T and TP taken together. It is hard to deny that the first and second steps are correct. However, the three steps are only as strong as the weakest, which is the third step. This step seems plausible, but is not undeniable. Still, I believe it is sufficiently strong to grant us our conclusion, that we have reason to believe

brain theory, provided that we also have reason to believe the other premises.

#### 9.4 The causal generalizing argument

Since its original formulation by Howard Robinson several different versions of the causal generalizing argument have been formulated.<sup>177</sup> Briefly, my version of the argument goes as follows. Take an arbitrary veridical perception; it has a corresponding possible hallucination\*, in which phenomenal qualities are instantiated in the brain of the subject. The instantiation of these phenomenal qualities in the hallucination\* is most proximately caused by a brain state. But, since the same brain state will be present in the veridical perception that corresponds to the hallucination\*, it follows by the principle "same sufficient proximate cause, same effect" that the same qualities should be instantiated in the brain also in the corresponding veridical perception.

Note that, although the causal generalizing argument does not use the type physicalism assumption, I used this assumption above, when arguing for the claim that hallucinations are instantiated in the brain in hallucinations.

However, an analogous version of the causal generalizing argument can also be used to generalize the conclusion that phenomenal qualities experienced in hallucinations are instantiated in sense data. Thus, independently of TP, we can use it to generalize the conclusion from ET and IT, that phenomenal qualities experienced in hallucinations are instantiated in internal objects, to veridical perceptions.

More formally, we may state the argument as follows:<sup>178</sup>

Let us define a hallucination\*\* as a hallucination\* where not only the brain states of its subject at its occurrence, but also the immediately preceding physical brain states of the subject, are the same as in its corresponding veridical perception. Using this notion we can formulate the thesis H\*\*T,

---

<sup>177</sup> See Robinson (1985) and (1994, chapter iv) for the original presentation of the causal generalizing argument and replies to some objections, including counterexamples to the "same proximate sufficient cause, same effect" principle. See also Martin (2004) and Johnston (2004) for further discussions about the argument.

<sup>178</sup> Note that in this argument I have made some steps, for which similar steps also occurred in the identity argument, more compact.

which is identical to H\*T, except that it concerns hallucinations\*\* rather than hallucinations\*.

P1 (H\*\*T) For each veridical experience there is a corresponding possible hallucination\*\*.

P2 (ET) We experience phenomenal qualities in hallucinations.

P3 The instantiation of phenomenal qualities in the brain in a hallucination is most proximately sufficiently caused by a physical brain state B.

P4 If the same proximate sufficient cause occurs, the same effect happens.

P5 (HBT) For all subjects S and phenomenal qualities Q: If S experiences Q in a hallucination, then Q is instantiated in the brain of S.

C1 (by H\*\*T) For an arbitrary veridical perception, V, where the phenomenal qualities Q<sub>1</sub>-Q<sub>n</sub> are experienced, there is a corresponding possible hallucination\*\*, H\*\*.

C2 (by C1, ET, HBT and definition of 'hallucination\*\*') Phenomenal qualities Q\*<sub>1</sub>-Q\*<sub>n</sub> are instantiated in the brain in H\*\* and these are indistinguishable from the phenomenal qualities Q<sub>1</sub>-Q<sub>n</sub> experienced in V.

C3 (by definition of hallucination\*\*) B occurs in V's subject's brain just prior to V.

C4 (by C3, P4) B causes the instantiation of phenomenal qualities Q\*<sub>1</sub>-Q\*<sub>n</sub> in S's brain in V.

C5 (by C2) Phenomenal qualities Q\*<sub>1</sub>-Q\*<sub>n</sub> are instantiated in the brain of S in V.

The final conclusion, C5 is equivalent to the conclusion of the identity argument, minus the claim that the Q\*s are physical properties. From this

thesis we can infer brain theory, using the type physicalism thesis and the arguments against the two sets view in 9.3 above.<sup>179</sup>

The conclusions follow from the premises. However, we have three new premises that have not been argued for, P1, P3 and P4. Let us look at them in turn briefly.

P1 is just a minor variation of H\*T. We can justify it using the super-scientist thought experiment, the scenario used to support H\*T, if we make a slight modification in the thought experiment. In the original scenario a subject first has a veridical perception. Then, her brain is kept in a constant state by a team of super scientists, as all external objects are being removed, resulting in her having a hallucination. We can modify the scenario by stipulating that the preceding brain state was the same in both the hallucination case and the veridical case. This makes the hallucination experience into an H\*\*. We have no reason to think that the H\* scenario is possible but the H\*\* is not. Thus, it seems that if we accept that there is a hallucination\* for each veridical perception, we should also grant that there is a hallucination\*\* for each veridical perception, and that P1 is justified.

P3 seems uncontroversial. The only alternatives to a physical brain state as the proximate cause of the instantiation of the qualities experienced in the hallucination are (i) direct physical causation from outside the brain, and (ii) that the event is uncaused. Here (ii) is implausible and (i) conflicts with our empirically informed models of perception, according to which the causal chain ending in the instantiation of the qualities Q1-Q<sub>n</sub> in H\*\* goes through the brain. Such "long-range" causation from an external event, that takes place without any mediating physical mechanisms, seems both implausible and contrary to our scientific worldview.

The weakest premise is P4. However, not even P4 is easy to deny. If the exact same physical brain state B occurs in two cases, how could the brain state not have the same effects? It seems that to deny P4 again "long-range" causation would be needed; in this case to inhibit the causal effects of B. But since we have no independent reason for believing in such long-range causation in perception, the postulation of it would be *ad hoc*, as Robinson (1994) points out. However, as I briefly consider next, the scope of the principle has to be restricted for it to be plausible.

---

<sup>179</sup> Note that without the type physicalism assumption it is not clear that we can use the arguments against the two sets view from section 9.3.

As I mentioned in the introduction to this chapter I will not be able to cover the full debate of the “same sufficient proximate cause, same effect” principle P4.<sup>180</sup> However, I will note the following objection to it. Harold Langsam (1997) has argued that there are counterexamples to the principle in P4 as it is stated. He agrees that the principle applies to intrinsic consequences but denies that it is applicable to *relational consequences*. We get an example of this inapplicability if we consider two almost identical cases of a worm moving across a table. The difference between the two cases is that in the first, a bottle stands at the edge of the table, whereas in the second there is no bottle. Now, it seems that although the intrinsic effects of the worm’s crawling are the same, the relational effects of the worm’s crawling differ. In the bottle case, one effect of the worm’s crawling is that the worm comes to stand in a certain spatial relation to the bottle, but this is not an effect of the worm’s crawling in the case where there is no bottle.

I will here merely note that even if we grant Langsam his conclusion, it does not threaten our application of the principle in this version of the causal generalizing argument. The reason is that the instantiation of phenomenal qualities in the brain is an intrinsic effect of the brain state that caused the instantiation, not a relational effect. Thus, even if we modify the principle P4 to only cover intrinsic effects, we still reach our desired conclusion.<sup>181</sup>

Since I have not been able to provide a full justification of P4 we cannot use the causal generalizing argument to support brain theory. However, we can note that if a suitably restricted version of P4 is true the argument can be used for this purpose.

Comparing the identity argument and causal generalizing argument, we see that the latter avoids the assumption of type physicalism in the generalization step, but does so at the cost of having to justify the same proximate sufficient cause, same effect principle. Thus, the causal generalizing argument can be used also by dualists, unlike the identity argument.

---

<sup>180</sup> Robinson (1994) replies to objections to the principle posed by Hinton, (1973) and Pitcher (1971).

<sup>181</sup> The reason Langsam (1997) argues for a modified principle is that he takes perception in the veridical case to be a relation. This helps against an argument saying that since no such relation is caused in the hallucination, no such relation is caused in the veridical case. But since our argument does not concern relational effects, but rather intrinsic effects, we need not worry about his modified principle.



# 10. Objections and replies

## 10.0 Introduction

In this chapter, I will reply to objections against brain theory, namely the “observation objection”, various common sense objections, the “extension objection”, the “epistemic objection” and the “phenomenal intransitivity argument”.

## 10.1 The observation objection

The observation objection starts from the observation that if we look at the exposed brain of a person who perceives something, such as when doing brain surgery on her, we will not ourselves experience the phenomenal colors and shapes she experiences. For example, if we look inside the brain of someone who experiences a red square, however closely we look, we will not experience a red square. But if phenomenal qualities are properties instantiated in the brain, as brain theory claims, we should be able to experience them, since we, in principle, can see the brain and all its parts. For example, if we are looking at the right places inside a brain of someone who experiences a red square, we should also experience a red square. So, the objection goes, the qualities cannot belong to the brain.<sup>182</sup>

We might express this objection more formally as follows:

*Implication:* Brain theory implies that we should always experience the phenomenal color and shape qualities which are instantiated in a brain we are looking at, if we look in the right places.

*Observation:* When looking inside a brain, we do not always experience the phenomenal color and shape qualities which, according to brain theory, are instantiated in it, even if we look in the right places.

---

<sup>182</sup> Smart (1959) addresses a similar objection to his identity theory of consciousness.

Thus, Brain theory is false. This objection might seem strong at first sight. However, it rests on an ambiguity, which gives rise to two different versions of the objection. When we disambiguate, we see that neither of these versions provide a good argument against brain theory.

As was discussed in Chapter 3, assuming brain theory is true, there are two different “perceptual” relations that we can stand in to things. If an object O causes an experience in a subject in the right way, she stands in the relation “perceiving” to O. On the other hand, if a subject stands to an object O in the more intimate relation that we stand in to the phenomenal qualities we experience, she stands in the relation “experiencing” to O.

Thus, we can distinguish between two readings of the term “experience” as it is used in the observation objection. This ambiguity gives rise to two distinct versions of the objection. First, the term “experience” might refer to the perceiving relation. Let us call this version the *perceive-version*. Second, the term “experience” in the objection might refer to the experiencing relation. Let us call this version the *experience-version*. I will look at both versions in turn and show that neither is a good argument against brain theory, beginning with the *perceive-version*:

*Implication<sub>p</sub>*: Brain theory implies that we should always *perceive* the phenomenal color and shape qualities, which are instantiated in a brain we are looking at, if we look in the right places.

*Observation<sub>p</sub>*: When looking inside a brain we do not always *perceive* the phenomenal color and shape qualities, which according to brain theory, are instantiated in it, even if we look in the right places.

Thus, Brain theory is false.

The *perceive-version* fails because of a dilemma. On brain theory, the phenomenal qualities we experience are identical to certain physical properties that are instantiated in our brains. Thus, when we are looking at the brain of another subject, who is experiencing phenomenal qualities, such properties are instantiated in her brain. Let us call these properties that are instantiated in the brain we are looking at the Ps.<sup>183</sup> When we are looking at

---

<sup>183</sup> These are neural correlates of experienced phenomenal qualities, in the sense that was discussed in Chapter 3: brain properties that are correlated with phenomenal qualities experienced by the subject of the brain, and are identical to the phenomenal qualities, according to brain theory.

her brain, we can either see the Ps or we cannot.<sup>184</sup> In either case the objection is unsound.

The first horn of the dilemma is that if we can see the Ps, *Observation<sub>P</sub>* is false. If we can see the Ps, in the sense that we can *perceive* them, the reason we can see them is that they cause experiences in us in the right way. However, by the indiscernibility of identicals, if certain phenomenal qualities, the Qs, are identical to these Ps, then it follows that the Qs also cause experiences in us in the right way. Therefore, we can perceive the Qs. Thus, *Observation<sub>P</sub>* is false in this case.

The other horn of the dilemma is that if we cannot see the Ps *Implication<sub>P</sub>* is false. The reason is that if the Ps are not visible, and the Qs are identical to the Ps, then the Qs are also not visible. But if the Qs are not visible we should not expect to be able to see them when we are looking at a brain in which they are instantiated even if brain theory is true. In other words, if the Ps cannot cause experiences in the right way in a subject which is looking at a brain in which they are instantiated, then it follows that neither can the Qs, since on brain theory they are identical with the Ps. Thus, if the Ps are not visible, brain theory does not imply that a subject looking at a brain should see phenomenal qualities that are instantiated in that brain, and *Implication<sub>P</sub>* is false. So, regardless of whether we can see the Ps or not, the perceive-version of the observation objection fails.

The experience-version of the objection goes as follows:

*Implication<sub>E</sub>*: Brain theory implies that we should always *experience* the phenomenal color and shape qualities which are instantiated in a brain we are looking at, if we look in the right places.

*Observation<sub>E</sub>*: When looking inside a brain we do not always *experience* the phenomenal color and shape qualities, which according to brain theory are instantiated in it, even in we look in the right places.

---

<sup>184</sup> Whether we can see the Ps depends on two things. First, naturally, it depends on what kind of properties the Ps are. But, second, it also depends on how much we can see. Philosophers disagree over which properties we can perceive. For example, historically, Berkeley argued that we could not perceive depth and Hume argued that we could not perceive causation. Recently, philosophers of mind have debated the closely related questions of which properties are included in the contents of perception and which properties we phenomenally perceive. An example of a relatively liberal theory of the contents of perception is Siegel (2006) according to which perception represents high-level properties like being a table or being a pine tree. An example of a conservative theory of phenomenal seeing is Price (2006). According to this we only phenomenally see color and position.

Thus, Brain theory is false.

The experience-version of the observation objection fails because *Implication<sub>E</sub>* is false. It does not follow from brain theory that we should always experience the same qualities as a brain we are looking at experiences, even if we look in the right places in the brain.

On brain theory, we stand in the experiencing relation to properties of our own brain, not to the properties of an external object we are looking at, regardless of whether or not that external object happens to be a brain. In other words, we stand in the perceiving relation to properties of external objects, which cause our experiences, and we stand in the experiencing relation to properties of our own brain. Thus, we should not expect to experience phenomenal qualities instantiated in other brains than our own brain, even if we are looking at these other brains.

We can see this more clearly if we take a closer look at brain theory's model of perception. As was described in Chapter 3, it states that the perceiving process proceeds roughly according to the following description. In perception, external objects reflect patterns of light beams which are detected by the eyes. Depending on their surface reflectance properties, the objects reflect different proportions of the incident light in different parts of the visible color spectrum. This information, along with perceptual clues, and perhaps other background knowledge, is used by the brain's visual system to generate internal representations of the external objects. The Ps, which the phenomenal qualities are identical to, enter at the last stage, since they are plausibly properties of such representations. Finally, there is some neural causal process that connects the Ps to whatever in our brain it is that constitutes our subject. This process grounds the experiencing relation, and so could aptly be called the *experiencing process*. Metaphorically speaking, we could think of this experiencing-process as mediated by an "inner eye," in contrast to the outer eye that mediates the perceiving-process in the case of seeing.

The important point to notice is that where the perceiving process connects us to external objects and their properties, the experiencing process connects us to properties of our own brain. Thus, we should not expect to experience phenomenal qualities instantiated in the brains of others, unless we have a similar neural connection to the Ps that are instantiated in these other brains as the connection we have to the Ps in our own brain.<sup>185</sup>

---

<sup>185</sup> A possible real life case of such a neural connection to a brain other than your own is that of Siamese twins conjoined at the head. However, just looking at the brain of someone else in the ordinary way, using your eyes, is clearly not enough.

So according to brain theory, the qualities we experience are instantiated in our own brain. We might be able to *see* qualities instantiated in other brains, in the perceiving sense of ‘see’, if we can see the Ps, but we cannot experience them. Thus, if brain theory is true, when looking at a brain we should not always directly experience the qualities that are instantiated in the brain even if we look in the right places, so *Implication<sub>E</sub>* is false.<sup>186</sup>

Why might one find *Implication<sub>E</sub>* compelling in the first place? The reason could be that it follows from a certain mix of the naïve view and brain theory. The naïve view, as defined in Chapter 2, could briefly be described as follows. Naively, that is, if we disregard our knowledge about the psychology and philosophy of perception, we take phenomenal qualities, such as phenomenal color and shape qualities, to be located out there in external objects, just as they seem to be phenomenologically.<sup>187</sup> What is special about the qualities you experience is that they are visible to you and being looked at. Conversely, when you look at a quality, such as an instance of phenomenal redness, which is instantiated in an external object, and the object is visible to you, you experience the quality. One might say that, according to the naive picture, vision is the process of external qualities just being out there in front of you becoming experienced by you when you look at them and the right conditions hold, the conditions sufficient for visibility.

Now we can see why *Implication<sub>E</sub>* is true, given a certain mix of the naïve view’s model of perception and brain theory. In the special case where a phenomenal quality is located in a part of a brain in front of us, and that part is visible to us, we should directly experience the quality if we look at the right part of the brain, according to the naïve view. If brain theory is taken to claim that the qualities of a brain we are looking at are located in said brain *Implication<sub>E</sub>* is true.

The mistake here consists in only applying brain theory to other perceivers than yourself, by taking the qualities they experience to be instantiated in their brains, rather than in the external objects they are looking at, while at the same time failing to apply brain theory to yourself, expecting to

---

<sup>186</sup> In the special case where you are looking at your own brain, perhaps through a mirror or a camera, you will naturally experience the qualities which are instantiated in the brain, but this will not be the case when you are looking at any other brain than your own. Even if the other brain happens to be looking at a brain which is indistinguishable from the brain you are looking at (perhaps itself), from the same direction you are looking at it from, you will not experience the qualities that are instantiated in the brain, although you will experience qualities indistinguishable from those that are instantiated in your own brain.

<sup>187</sup> Strictly speaking, I have only claimed that this claim is included in the naïve view of perception, when discussing it in Chapter 1. However, plausibly, the following claims are also included in our naïve view of perception, and if they are included they can explain why one might find *Implication<sub>E</sub>* compelling.

experience qualities instantiated in the things you look at rather than the qualities instantiated in your brain.<sup>188</sup>

## 10.2 The common sense objection

One general type of response to brain theory is to say that it is plainly false, absurd or mad to claim that the phenomenal qualities belong to the brain. Such crude responses could be refined into objections such as that brain theory is contrary to common sense, counterintuitive or against our pre-theoretical beliefs, and that because of this we should not believe it unless we have strong arguments speaking in favor of it.

I will distinguish three distinct common sense arguments, resulting from three different things one might mean by the phrase “common sense.” They all have the same general structure, claiming that brain theory is contrary to common sense in a certain meaning of the phrase, and that because of this we should not believe it.

By “common sense” one might have in mind what is expressed by statements people typically accept.<sup>189</sup> For example, we commonly accept statements like “Hands ordinarily have five fingers.” or “Tables are often made of wood.” so the contents of these statements might be considered common sense.

For two reasons, it is far from obvious that brain theory conflicts with the contents of statements people typically accept. First, it is unclear whether any typical statements cover the claims of brain theory. It might be the case that people generally do not accept statements like “the brain is red and square.” But it is less certain that they thereby deny what brain theory claims, that the brain instantiates properties like phenomenal redness and phenomenal squareness. Keep in mind that brain theory is neutral on whether the properties denoted by sensible properties terms belong to the brain. Phenomenal qualities are denoted by technical terms like “phenomenal redness” and “phenomenal squareness”, rather than ordinary

---

<sup>188</sup> A possible explanation for why people make this mistake is that everyone is familiar with the naïve view, whereas brain theory is a new and unfamiliar theory.

<sup>189</sup> Because of the emphasis on ordinary language, this first sense is related to the philosophical tradition emerging from Thomas Reid and the Scottish school of common sense. According to Reid common sense constitutes the set of beliefs which we cannot help believing because of our human nature. He suggested that we can find out about these by studying ordinary language, in particular, structures common to all languages. The reason for this was that he saw ordinary language as closely connected to our ordinary everyday thoughts, which were supposed to depend upon the dictates of human nature (Yaffe and Nichols 2009).

words like “red” or “square.” Since people, in general, typically do not use such technical terms, we have reason to believe their typical statements are silent on the metaphysical location of the qualities.

Second, it is not even that clear that people typically do not accept statements such as “the brain is red and square,” if the context is a discussion about perception. As was discussed in the introduction many scientists state things like that when talking about perception. Arguably, brain-theory-like views of perception are even common among scientists. Since it is hard to know how much scientists have affected the general population we cannot be certain that such statements are not also widespread in the general population until we have investigated the question empirically.

Even if brain theory were contrary to statements we typically accept, this would not provide a significant reason against it. Initially, one might wonder why it should be relevant at all in an abstract and complex matter like this, if people generally state things that are or are not compatible with this view. But even if we assume that it is sometimes relevant that there are particular reasons to believe, it is not relevant in this case. We will see this after we have looked at typical cases where it is not relevant.

Many true statements, such as huge parts of physics, at least at first sight go against things we ordinarily say. Among the most well-known are statements like “the sun rises” which taken at face value seems to imply that the sun moves relative to a stationary earth. Assuming the statement expresses its face value content, this statement is contrary to the well-known astronomical result that the earth revolves around the sun, rather than the other way around. But nevertheless, we should believe the astronomical result and not the common sense statement. Here we, thus, have an example of a case where we should trust science even if it goes against the things we typically say.

Plausibly, the reason science trumps typical common sense statements, is that we have good independent reasons for believing in scientific statements, possessing evidence in their favor. It thus seems like our typical statements can, at best, provide weak reasons which only affect whether we should believe a view if all else is equal when it comes to evidence for it. Since, as I argue in the rest of my thesis, all else is not equal when it comes to brain theory, we should not disbelieve it even if it goes against what people typically say. Thus, the first common sense argument does not give us any reasons against believing in brain theory.

Let us proceed to the second common sense objection. Another thing one might mean by the phrase “common sense” is the beliefs contained in the

*manifest image*.<sup>190</sup> This concept stands for the sum of our naïve everyday thoughts about the world. For example, it is part of the manifest image that external objects are phenomenally colored and that typical objects like tables are solid through and through. Certain scientific claims, such as that an ordinary table is mostly composed of empty space, on the other hand seem to go against manifest image.

Just as the first meaning of "common sense," this second meaning also concerns the way people ordinarily think about things. However, there are some differences. First, the preceding meaning focuses explicitly on language whereas the second focuses on beliefs. Second, and more importantly, the beliefs in the manifest image are pre-theoretical. Thus, as long as people are typically pre-theoretical on metaphysics of perception, the first and the second common sense objections coincide, but they will be distinct otherwise.

Unlike on the first meaning of common sense, it is relatively clear that brain theory is contrary to the manifest image. Our naïve view of perception includes that qualities are located in external objects rather than in brains, at least in veridical perceptions. The core identity claim of brain theory denies this.

Still, for the same reasons as those that were raised against the first objection, a conflict with the manifest image does not provide significant reasons against brain theory. There are counterexamples, such as many scientific claims, which *prima facie* are in conflict with the manifest image, but which nevertheless should not be abandoned. The reason we should not abandon these scientific claims, despite the conflict, is that we have independent evidence for them. Furthermore, even if it turns out that the naïve claims that *prima facie* seem to be incompatible with science are not really incompatible with science, this does not matter. The important point is that *if* there is such an incompatibility, we should believe in the evidence-based scientific claims rather than the conflicting claims from the manifest image. Since, as I argue in the rest of this thesis, the weight of the independent evidence favors brain theory, we should not abandon it, even if it conflicts with the manifest image.

---

<sup>190</sup> The manifest image is closely related to our naïve view of the world, as was discussed in Chapter 1. There the relation to how the world seems through phenomenological reflection was also described. Because of this I will treat objections against brain theory from how the world naively seems or from how it seems through phenomenological introspection as coinciding with this objection.

Let us proceed to the third common sense objection. On a third interpretation of “common sense” the phrase refers to sound judgment.<sup>191</sup> Extending this sense, we might say that beliefs formed through sound judgment are common sense. For example, if someone has heard good evidence for a claim and knows of no evidence against it, then it would be common sense to believe the claim and the belief could be called common sense.

Brain theory does not go against common sense, in this third meaning of ‘common sense’. If the evidence favors a theory, so does sound judgment, and vice versa. But given that the rest of my thesis is roughly correct, the evidence actually favors brain theory, as does sound judgment. If, on the other hand, brain theory had been contrary to sound judgment, we should not believe it. However, this is not so, assuming that my arguments for brain theory presented in previous chapters, as well as the other replies to objections present in this chapter are correct. Thus, the third common sense objection does not provide any reasons against brain theory.

### 10.3 The extension objection

The idea behind the next objection is that some phenomenal qualities, such as phenomenal color qualities, necessarily are properties of extended objects, if they are properties of anything. But brain states belong to the ontological category of states and states are not spatially extended. Thus, phenomenal colors cannot belong to brain states.

---

<sup>191</sup> Compare the following dictionary definitions of common sense:

“sound and prudent judgment based on a simple perception of the situation or facts” (Merriam-Webster)

“the ability to think and behave in a reasonable way and to make good decisions ” (Merriam-Webster learner’s dictionary)

“the basic level of practical knowledge and judgment that we all need to help us live in a reasonable and safe way” (Cambridge dictionary)

“the ability to think about things in a practical way and make sensible decisions” (Oxford advanced learner’s dictionary)

Perhaps the most central common theme here is that common sense is related to reasonable actions, prudence, and making good decisions. However, such properties of persons do not directly transfer to properties of beliefs. A second theme, closely related to sound judgment does transfer. We thus get our third sense, common sense as beliefs based on sound judgment.

Note that this objection only applies to *one version* of brain theory. As was stated in section 3.1, brain theory is neutral on what in the brain phenomenal qualities are instantiated in. They could be instantiated in the brain, in some part of the brain or in brain states. Since both the brain and parts of the brain are clearly spatially extended, whereas brain states, according to the objection, are not spatially extended, this objection is plausibly only an objection against the version of brain theory which claims that phenomenal qualities are properties of brain states. Let us call this version *brain state theory*. I will argue that even this version of brain theory can avoid the objection.

A version of the extension objection has been used by Shoemaker (1996b) in an argument against “literal projectivism”:<sup>192</sup>

I am looking at a book with a shiny red cover. The property I experience its surface as having, when I see it to be red, is one that I can only conceive of as belonging to things that are spatially extended. How could that property belong to an experience or sensation? Remember that an experience is an experiencing, an entity that is 'adjectival on' a subject of experience. It seems no more intelligible to suppose that a property of such an entity is experienced as a property of extended materials than it is to suppose that a property of number, such as being prime or being even, is experienced as a property of material things. The literal projectivist view may seem more palatable if the projected properties are said to be properties of portions of the visual field (see Boghossian and Velleman 1989). But that, if taken literally, amounts to a resurrection of the sense-datum theory, with all its difficulties. (Shoemaker, 1996b, 250-251)<sup>193</sup>

Shoemaker claims that he can only conceive of color qualities as properties of things that are spatially extended. Brain state theory claims that qualities are properties of brain states; and brain states belong to the metaphysical category *states*. Plausibly, states are a subcategory of the metaphysical category of *events*.<sup>194</sup> Thus, a brain state is an event. So if events lack extension then Shoemaker’s objection is not only an objection against literal projectivism, but also an argument against brain state theory. Although it is far from clear that events actually lack extension, let us assume this for the

---

<sup>192</sup> Literal projectivism is the view that experiences systematically misrepresent features of experiences as belonging to the external environment. Since brain theory does not claim that phenomenal qualities are features of experiences, it is not literally a version of literal projectivism. However, brain theory does claim that phenomenal qualities are internal, in the sense that they belong to the brain. If our experiences represent phenomenal qualities as belonging to external objects, as might be plausible, then brain theory is, at the very least, closely related to literal projectivism.

<sup>193</sup> See also Shoemaker (1996b, 102)

<sup>194</sup> See Casati and Varzi (2011).

sake of argument<sup>195</sup>. We may express the extension objection in terms of premises and conclusion, as follows:

P1 Necessarily, if phenomenal color qualities are properties of anything, they are properties of spatially extended entities.

P2 Brain states are not spatially extended

Thus, these qualities are not properties of brain states and brain state theory is false.

It is possible to deny the first premise. Such a denial could be supported by arguing that our intuitions about things like this are not a good source of knowledge. On the other hand some things seem to speak in favor P1. For instance, it might be hard to conceive how experienced redness could be a property of an extensionless point, and even harder to conceive how it could be a property of something completely non-spatial. Thus, because P1 is not clearly wrong, I will not push this way of replying to the extension objection.

However, regardless of whether or not color qualities must be properties of something extended, the objection fails. The reason is that the objection rests on an ambiguity. On brain theory, we should make a distinction between physical spatial properties and phenomenal spatial properties. The spatial properties we *experience*, such as the squareness of a hallucinated red square, are phenomenal spatial properties. On the other hand the physical shapes that external objects, including brains, have are physical spatial properties.

We can thus read the objection in two ways: either as concerning physical spatial properties or as concerning phenomenal spatial properties. As we shall see, both versions fail. The first version goes as follows:

P<sub>1P</sub> Necessarily, if phenomenal color qualities are properties of anything, they are properties of physically spatially extended entities.

---

<sup>195</sup> There are different metaphysical theories of events. According to some, for example Quine (1960), there are no relevant differences between objects and events; both are four-dimensional entities that extend in both time and space. Like objects, events are also both spatially and temporally extended, although events tend to take up less time than objects (Casati and Varzi, 2011). On these theories, plausibly brain states would be four-dimensional entities extended throughout the parts of the brain they belong to. Thus, brain states would be extended, and the extension objection fails. According to other metaphysical theories of events, such as Chisholm's (1970) theory, events are more like facts. If one of the latter theories is true, events are unextended and it is correct that brain states are not spatially extended.

P<sub>2P</sub> Brain states are not physically spatially extended.

Thus, these qualities are not properties of brain states, and brain state theory is false.

The truth of P<sub>2P</sub> follows from the assumption that events are not physically spatially extended, since brain states are events. However, regardless of whether introspective intuitions can provide evidence or not, we have no reason to believe P<sub>1P</sub>. The reason for this is that the intuitions that lie behind the extension objection are, at least partially, based on introspection. For example, when stating the intuition behind the extension objection, Shoemaker claims that the phenomenal qualities are *experienced as* belonging to external objects. Thus, the underlying intuitions plausibly concern phenomenal extension, rather than physical extension, since it is the former that we experience. But as P<sub>1P</sub> concerns physical extension, and not phenomenal extension, these intuitions give no support to P<sub>1P</sub>. Since we do not have a good reason to believe P<sub>1P</sub> anymore, the first version of the argument fails.<sup>196</sup>

The second version is also unsound. It goes as follows:

P<sub>1E</sub> Necessarily, if phenomenal color qualities are properties of anything, they are properties of phenomenally spatially extended entities.

P<sub>2E</sub> Brain states are not phenomenally spatially extended.

Thus, these qualities are not properties of brain states, and brain state theory is false.

This time, assuming that introspective intuitions can provide evidence, the intuitions that underlie the extension objection provide evidence for P<sub>1E</sub>, since this premise concerns phenomenal extension. However, we have no reason to believe P<sub>2E</sub>. This is because the fact that brain states are not physically extended is compatible with brain states being *phenomenally* extended, and the fact that we have no other reason to believe that brain

---

<sup>196</sup> Note that P<sub>1P</sub> might still be true, even if we do not have intuition based evidence for it. If another theory of phenomenal qualities than brain state theory is true, and that other theory claims that phenomenal qualities belong to physically spatially extended entities, such as the brain itself for instance, then plausibly the phenomenal qualities are *necessarily* properties of something physically spatially extended. However, this possibility cannot be used to argue against brain state theory, since the assumption that another theory of phenomenal qualities is true presupposes that brain state theory is false, something that an argument against brain state theory cannot do without begging the question.

states are not phenomenally extended.<sup>197</sup> Thus, the second version of the extension objection also fails.

#### 10.4 The epistemic objection

On one of the three criteria of directness discussed in 3.3, experience-directness, brain theory is an indirect theory of perception. Some philosophers have claimed that indirect theories of perception face insurmountable epistemological problems.<sup>198</sup> We might formulate the standard objection as follows.<sup>199</sup> If we are never directly aware of external objects, but instead ever only directly aware of some internal intermediates, such as sense data, then we cannot justify our belief in an external physical world. A natural reply is that an inference to the best explanation, from the ordered structure of our experiences, can justify belief in an external physical world. But, argues the direct realist, contrary to what philosophers such as Russell (1997, 22-4) have claimed, the attempt to justify our belief in an external world through such an inference is a hopeless project. Furthermore, it is not just our general belief in the existence of an external world that is threatened, but also our beliefs about particular facts about the external world, such as my belief that I am sitting on a chair, in front of a table, while writing this. On the other hand, the objection assumes, a direct realist, who claims that in veridical perceptions we are directly aware of external objects, can avoid these skeptical problems, or at least has an easier time dealing with them.

There are two ways to respond to the epistemic objection. First, one might defend the project of justifying our beliefs about the external world through an inference to the best explanation against the objections which have been raised against the project. This has been done by Jackson (1977) and Robinson (1994). Second, one might argue that, regardless of whether an inference-to-the-best-explanation strategy works or not, it is not obvious

---

<sup>197</sup> Brain state theory even claims that brain states are what instantiate phenomenal extension and other spatial phenomenal qualities, since the general version of brain theory claims that these qualities are properties of the brain.

<sup>198</sup> According to both Robinson (1994, 215), Jackson (1977, 138), Brown (1992, 341) and Ward (1976, 287) the epistemic objection is among the most prominent objections against indirect realism, if not the most prominent objection. An example of a philosopher who employs it is Cornman (1975). Berkeley (1710) made an early epistemic argument against indirect theories.

<sup>199</sup> See Brown (1992) and Robinson (1994, 215-217) for descriptions of the epistemological objection to indirect perception along these lines.

that direct realism is in a better position than indirect realism when it comes to justifying our beliefs about the external world. I will present a reply of that second type below.<sup>200</sup>

Contrary to a common presumption, it is not obvious that indirect theories of perception face worse skeptical challenges than direct theories. One way to illustrate this is to distinguish between two types of views on justification, *internalist* and *externalist* theories of epistemic justification. Below I will argue that it is not clear, on either view, that direct realism has advantages over indirect realism, when it comes to perceptual justifications of our beliefs about external objects.

According to internalism about justification, for a belief to be justified its subject must have some accessible basis for the belief.<sup>201</sup> For example, my belief that there is a tree in front of me can be justified if I have an appropriate visual experience of that tree, with the content that there is a tree in front of me.

For direct realism to have an advantage over indirect realism in terms of internalist justification, the truth of direct realism must imply that perceivers have some additional accessible basis for their perceptual beliefs, which perceivers lack if indirect realism is true. But it is not obvious that direct realism can provide such a basis.

Note first that veridical experiences have the same phenomenal character on both theories, and that experiences can even have the same contents on both theories. So, on direct realism neither the phenomenal character nor the content of experiences can provide us with an internalist justification for belief in external objects, that the content or character cannot provide us with on indirect realism.

Furthermore, direct realists concede that it is possible to have hallucinations that are indistinguishable from veridical perceptions. Thus, as BonJour (2012) argues, on direct realism we seem to have the same reasons for belief in external objects in both the bad and the good case. But if we have the same reasons in these two cases, which differ in that we are directly aware of external objects in one case but not in the other, then we should have no better reasons for believing in an external world in the good case, conditional on direct realism being true, than we should have in the good case if indirect realism is true.

---

<sup>200</sup> I am indebted to Fiona McPherson for the line of thought behind this reply.

<sup>201</sup> See BonJour (2012).

Thus, it is not obvious that direct realism has an advantage over indirect realism, on internalism. BonJour (2012) draws a similar, although stronger, conclusion (note that he uses the term “representative realism” instead of “indirect realism”):

The conclusion that seems indicated is that the very idea that direct realism represents a further alternative on the present issue is a chimera: that there is no sense in which the experience of material objects is plausibly direct that helps at all with the epistemological issue of justification. Thus, once phenomenalism is rejected as hopeless, the only alternatives with regard to knowledge of the external world appear to be skepticism and some version of representative realism. A fundamentally representative realist view can incorporate the view that the perceptual awareness of material objects is direct in the unproblematic, but epistemologically irrelevant ways noted earlier - that it is not arrived at via any sort of inference and that it intuitively “presents” material objects to the perceiver. It can even incorporate some version of disjunctivism, should one be defensible. But these refinements, valuable as they may be for other kinds of reasons, will apparently make no real difference as regards the fundamental epistemological issue (BonJour 2012).<sup>202</sup>

Externalist theories of epistemic justification deny that, to be justified, beliefs must have some basis that is accessible to the believer. Instead, on externalism, whether or not a perceptual belief is justified can be determined by factors that are not accessible to the believer. For example, in the case of a perceptual belief, facts about the process which caused the belief can determine whether the belief is justified or not.

A standard example of externalism is reliabilism.<sup>203</sup> According to this theory, whether a perceptual belief is justified or not depends entirely on if the belief was produced by a reliable process. Here, a process is taken to be reliable if it produces a sufficiently high proportion of true beliefs. For example, if I see a tree, and because of this form the belief that there is a tree in front of me, that belief is justified because my visual system is a system that produces a sufficiently high proportion of true visual beliefs. Note that on reliabilism it does not matter whether the subject of the belief *knows* that the process which caused her belief is reliable or not, what matters is that the process *is* reliable. Thus, for example, I do not have to know anything about my visual

---

<sup>202</sup> O'Brien (2013) draws the same conclusion. See also Ward (1976) for another criticism of the idea that a direct realist can give an internalist justification for belief in an external world, without using an inference to the best explanation from the structure of our experiences.

<sup>203</sup> For a classic version of reliabilism, see Goldman (1986).

system, or even that I have one, for my belief about the tree to be justified, in the case above.<sup>204</sup>

On externalism, it is enough for our perceptual systems to actually be sufficiently reliable for us to have justified perceptual beliefs about external objects. This justification is independent of the dispute between direct and indirect theories of perception, since that dispute does not concern the empirical facts that underlie whether our perceptual systems are reliable or not. If our perceptual systems are unreliable then we lack perceptual knowledge about external objects on both direct realism and indirect realism. On the other hand, if our perceptual systems are reliable then we have perceptual knowledge about external objects on both direct and indirect realism, regardless of whether the inference to the best explanation approach to justifying our belief in an external world works or not. Thus, on externalism direct realism does not have any obvious advantage over indirect realism for justifying our belief in an external world.

Let me sum up the conclusions of my discussion of the epistemic objection. First, brain theory is only an indirect theory on one of the three criteria for direct perception I discussed in 3.3. Second, indirect theories can justify beliefs about external objects through an inference to the best explanation. This strategy has been criticized, but there are also replies to the criticism. Third, even if brain theory is an indirect theory and the criticism against the inference to the best explanation strategy is successful it is still not obvious that the epistemological objection succeeds. The reason is that, regardless of whether internalism or externalism about justification is true, direct theories do not obviously have an easier time justifying our belief in an external world

## **10.5 The phenomenal intransitivity argument**

A common objection to sense data theory is that we know from sense experience that phenomenal indistinguishability is intransitive, and that this alleged fact is incompatible with sense data theory. Since brain theory has some similarities to sense data theory, we might wonder whether this is a problem for brain theory, as well. In this section, I will first briefly look at how the phenomenal intransitivity argument can be used against sense data

---

<sup>204</sup> See BonJour (2012).

theory, and then sketch a parallel argument aimed at brain theory. Finally, I will present two replies to the argument against brain theory.

D.M. Armstrong (1968) gives a classic exposition of the alleged intransitivity phenomenon, and uses it in an objection against sense data theory:

‘Exactly similarity in a particular respect’ is necessarily a transitive relation. Now suppose that we have three samples of cloth, A, B and C, which are exactly alike, except that they slightly differ in color. Suppose further, however, that A and B are perceptually completely indistinguishable in respect of color, and B and C are perceptually completely indistinguishable in respect of color. Suppose, however, that A and C can be perceptually distinguished from each other in this respect. Now consider the situation if we hold a ‘sensory item’ view of perception. If the pieces of cloth A and B are perceptually indistinguishable in color, it will seem to follow that the two sensory items  $A_1$  and  $B_1$  that we have when we look at the two pieces actually are identical in color. For the sensory items are what are supposed to make a perception the perception it is, and here, by hypothesis, the perceptions are identical. In the same way  $B_1$  and  $C_1$  will be sensory items that are identical in color. Yet, by hypothesis, sensory items  $A_1$  and  $C_1$  are not identical in color! (218)<sup>205</sup>

The “sensory items” Armstrong discusses are sense data. Since we get a contradiction, the assumption that there are sense data must be false, according to the argument.

There are various ways to respond to this argument.<sup>206</sup> First, one might deny that indistinguishability is an intransitive relation.<sup>207</sup> Second, one might argue that there can be vague objects, which can have indeterminate properties, and that sense data are such vague objects.<sup>208</sup> Third, one might deny that sense data necessarily are as they appear.<sup>209</sup>

A parallel argument phrased in terms of phenomenal qualities would say that if indistinguishability is an intransitive relation there can be cases, such as the three pieces of cloth Armstrong describes, where the members of two of the pairs cannot be distinguished from each other by a perceiver, but the members of the third pair can be distinguished from each other. If phenomenal qualities are experienced in perception, assuming phenomenal

---

<sup>205</sup> Nelson Goodman (1951) gives a famous early statement of the idea that indistinguishability is an intransitive relation.

<sup>206</sup> For discussion of indeterminacy objections against sense data, see Huemer (2001) and Huemer (2011).

<sup>207</sup> This route is explored by Robinson (1994, 190-191) and taken by Graff (2001) as well as Jackson and Pinkerton (1973).

<sup>208</sup> Cf. Ayer (1963). For a classic argument against the possibility of vague objects, see Evans (1978).

<sup>209</sup> Cf. Robinson (1994, 191-192).

qualities are as they appear to be, the three pieces of cloth Armstrong describes in the quote above, when perceived, give rise to a series of phenomenal qualities, which seem to have phenomenal color properties that are contradictory in the same way as the sense data  $A_1$ ,  $B_1$  and  $C_1$  had contradictory color properties. So, if we assume that phenomenal qualities are experienced in perception, and can be instantiated, we get a contradiction. Thus, there cannot be any phenomenal qualities, and brain theory is false.<sup>210</sup>

The same three types of replies that a sense data theorist could use against the intransitivity argument against sense data can also be used against this intransitivity argument against phenomenal qualities. However, I shall neither defend the claim that indiscriminability is transitive nor that there can be vague entities. Rather, I will give two other replies. First, the intransitivity objection is not an objection against brain theory primarily, but rather a problem for all theories that are realist about phenomenal qualities, in the sense that they claim that phenomenal qualities can be instantiated in; and this is claimed by most, if not all, theories of phenomenal qualities we are considering. Thus, even if the argument is correct, its conclusion is not a problem for brain theory in particular. Second, I will argue that the objection can be avoided, since phenomenal qualities are not necessarily as they appear. It does not follow that two phenomenal qualities must be qualitatively identical, just because we cannot distinguish them. Furthermore, this is not as strange as it may seem on first sight, since we have independent reasons to believe that we are sometimes mistaken about our phenomenology.

Let us look at these replies in turn. The first is straightforward. The original conclusion of the intransitivity argument against sense data is that there are no sense data. Similarly, the conclusion of the intransitivity argument against phenomenal qualities is that there are no phenomenal qualities. The argument does not just deny that experienced phenomenal qualities are instantiated in the brain of the subject of a hallucination, or that experienced phenomenal qualities are ever instantiated in the brain of a subject. What the objection concludes is that there are never *any* phenomenal qualities *anywhere*, including in veridical perceptions.

Of all the alternatives to brain theory that are compatible with the realism assumption from Chapter 5, only NN-theories, which claim that phenomenal qualities are neither instantiated in the good case nor in the bad case, deny

---

<sup>210</sup> Everett (1996) has given an argument of this type. He argues that the alleged phenomenon of phenomenal intransitivity implies that qualia are vague entities, and that vague entities are necessarily non-existing. See also Dummett (1975).

that phenomenal qualities are instantiated in veridical perceptions. Thus, among the alternatives we are considering, the intransitivity objection, even if correct, could only give NN-theories an edge over brain theory.

It might even be argued that the intransitivity argument has a stronger conclusion than this, namely that we cannot *experience* phenomenal qualities.<sup>211</sup> We can sketch such an argument as follows. If we have reason to believe that phenomenal qualities are as they appear, we should also have reason to believe that experiences of phenomenal qualities are as they appear. However, then it follows, by the intransitivity argument, that there is a series of experiences, with contradictory experienced phenomenal color properties. Thus, experiences of phenomenal qualities themselves must be indeterminate entities. But since there cannot be any indeterminate entities there cannot be any experiences of phenomenal qualities.

Not only is the conclusion of this stronger intransitivity argument equally a problem for all the theories of phenomenal qualities we are considering, but it also directly contradicts the realism assumption from Chapter 5, which says that:

*Realism assumption:* When a subject S has a veridical perception, S experiences phenomenal qualities.

This first reply to the intransitivity argument is not conclusive, since I have only sketched the part of the reply claiming that the intransitivity argument is a problem for all theories we are considering, including NN-theories. But regardless of whether the stronger conclusion follows or not, the reply shows that even if the intransitivity argument is correct the consequences for brain theory are far from devastating, since only the relatively rare NN-theories can benefit from it.

Let us proceed to the second reply. As Armstrong himself notes, one way to avoid the intransitivity objection is to “abandon the view that we have incorrigible knowledge of the nature of the items at the time of having them (Armstrong, 1968, 219)”.<sup>212</sup> This abandonment allows us to claim that phenomenal qualities may have finer details than our introspection can

---

<sup>211</sup> This is what Everett (1996) claims. He uses the intransitivity objection to argue, not only that there are no qualia, but also that we do not experience qualia. Plausibly, his qualia refer to phenomenal qualities. Graff (2001) has also pointed out that the intransitivity of indistinguishability problem is a general problem for all realists about phenomenal qualities. Furthermore, Deutsch (2005) has specifically argued that many versions of intentionalism cannot handle the intransitivity paradox, in particular, the strong reductive versions of it.

<sup>212</sup> As noted in Huemer (2011) this reply might be unappealing to traditional sense data theorists, since many of the motivations for sense data theory may be incompatible with denying that sense data have exactly the properties they appear to have. However, as brain theory does not have any motivations incompatible with this claim it is free for us to make this reply.

distinguish. Thus, even if indistinguishability is an intransitive relation and there are three pieces of cloth, that are such as Armstrong describes when talking about A, B and C in the quote above, the three instantiations of experienced phenomenal color qualities that correspond to these pieces of cloth, when the pieces are perceived, need not have contradictory phenomenal color properties. Rather, it could be the case that two of them are of the same phenomenal color, whereas the third instantiation is of a distinct, but very similar, phenomenal color. Thus, phenomenal qualities need not be indeterminate objects in cases like this, even if indistinguishability is an intransitive relation, and the intransitivity objection is avoided.

Is it reasonable to deny the principle that we always have incorrigible knowledge about the phenomenal qualities we experience? I believe we have at least two independent reasons for denying this principle. First, when we make judgments about phenomenal qualities, a mechanism is involved. But it should be possible for any mechanism to go wrong, especially in such a demanding case as when trying to determine whether or not two barely noticeable color qualities are exactly the same type of quality, or just extremely similar qualities. Second, results in experimental psychology indicate that we do not have full knowledge about our own phenomenology. In experiments on change blindness and inattention blindness, a test subject is first presented with a picture and instructed to look for changes in the picture. Unknown to the subject the picture is then instantly altered by the psychologists.<sup>213</sup> It can take subjects many seconds, or even minutes, to detect that the picture has changed, even if the change is substantial. However, presumably the qualities the subjects experience when looking at the picture, changed very soon after the picture was altered. As the subject would say that her experience has not changed, although it had changed, we here have a case where a subject is plausibly wrong about the phenomenal qualities she experiences. Thus, this case falsifies the principle that we always have incorrigible knowledge about the phenomenal qualities we experience.<sup>214</sup>

I conclude that the second reply provides an adequate response to the phenomenal intransitivity objection against brain theory. This conclusion finishes the Objections of replies chapter. As we have seen, brain theory has been able to respond to all objections against it, including the observation

---

<sup>213</sup> This alteration takes places during a saccade of the eyes, in a place of the picture the pupils are not directed at.

<sup>214</sup> See also the fascinating research conducted by Eric Schwitzgebel (for example in Schwitzgebel 2007). He has documented cases, ranging from dreams to what the pressure under our feet feels like, where we seem to lack perfect knowledge of our own experiences.

objection, three common sense objections, the extension objection, the epistemic objection and the phenomenal intransitivity argument.



## 11. Brief summary of the results

My argumentation has two main conclusions. The first is that we have reason to believe the experience thesis (ET) and the instantiation thesis (IT). I have argued in Chapter 6 that theories which deny the experience thesis, namely EX theories, face serious problems. In Chapter 7 I similarly argued that theories which deny the instantiation thesis, namely EN and NN theories, face serious problems. In Chapter 10 I argued that at least one theory which accepts the experience thesis and the instantiation thesis, brain theory, does not face serious problems. Together, these conclusions give us reason to accept the experience thesis and the instantiation thesis, which together state that we experience phenomenal qualities in hallucinations, and that these qualities are instantiated in existing internal objects, such as brains or sense data. The type physicalism assumption was not used in these chapters, so the conclusion that we have reason to accept both the experience thesis and the instantiation thesis is independent of it. The most important part of this result is that we have reason to accept the instantiation thesis, which is a more controversial thesis than the experience thesis.

The second conclusion is that we have reason to believe brain theory. I argued in Chapter 9 for the conditional claim that brain theory can be reasonably inferred from the instantiation thesis, the hallucination\* thesis (H\*T), the experience thesis and type physicalism, taken together. I defended the hallucination\* thesis in Chapter 5, using the super scientist thought experiment, and we have assumed type physicalism. Thus, if these steps are correct, since we have reason to believe the experience thesis and the instantiation thesis it also follows that we have reason to believe brain theory.



## Acknowledgements

This work was created through stimulating, sometimes intense, dialogue with my advisor Pär Sundström, often over a good lunch. I am deeply grateful for all his sharp and well thought-out comments and advice, as well as for his strong encouragement, especially during the final months.

My assistant advisor, Sten Lindström, contributed not only by giving comments, but through long hours of late night philosophical discussions, as well as by his incredible sense of humor and logical point of view.

Large parts of this dissertation have been presented and discussed at the "Friday seminars" in Umeå. I am grateful to everyone who has participated. In particular, I want to thank the "regulars" Anders Berglund, Emma Beckman, Petra Green, Ebba Gullberg, Peter Karlsson, Jonas Nilsson, Inge-Bert Täljedal, Lars Samuelsson, Bertil Strömberg and Mats Wahlberg, who have contributed with many useful comments and discussions. Furthermore, I would also like to thank everyone who has read and commented on previous versions of this manuscript, in particular Geoff Anders, Oscar Horta, Leonard Katz, Pablo Stafforini and Aron Vallinder.

But life is not merely work. Thus, I also want to thank my dear friends in Umeå, in particular Anna Bohman, Johannes Frech, Pär Sundling and the members of "the cult", among many others, for bringing warmth in the cold northern climate. Last but not least I want to thank my family: my parents Gerd and Stig-Björn and my grandmother Inez, who have been a constant source of support and to whom I have dedicated this dissertation.

Jesper Östman  
Göteborg, May 2013



## References

Alston, W. P., 1967, "Pleasure", in *The Encyclopedia of Philosophy*, P. Edwards (ed.), London and New York: Macmillan, Vol. 6:341-347.

---, 1999, "Back to the theory of appearing", *Philosophical Perspectives*, 13 (s13):181--203.

Alter, T. and S. Walter, (eds.), 2007, *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, New York: Oxford University Press.

Armstrong, D.M., 1968, *A Materialist Theory of the Mind*, London: Routledge.

---, 1989, *Universals: An Opinionated Introduction*, Boulder: Westview Press.

---, 1999, "The Open Door," in H. Sankey, (ed.), *Causation and Laws of Nature*. Dordrecht: Kluwer Academic Publishers, 175-85.

Aydede, M., 2013, "Pain", *The Stanford Encyclopedia of Philosophy (Spring 2013 Edition)*, Edward N. Zalta (ed.), forthcoming URL = <<http://plato.stanford.edu/archives/spr2013/entries/pain/>>.

Ayer, A. J., 1963, *The Foundations of Empirical Knowledge*, London: Macmillan.

Backhaus, W. and R. Menzel, 1992, "Conclusions from color-vision of insects", *Behavioral and Brain Sciences*, 15(1):28-30.

Bacon, J., 1995. *Universals and Property Instances: the Alphabet of Being*, Oxford: Blackwell.

---, 2011, "Tropes", *The Stanford Encyclopedia of Philosophy (Winter 2011 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/win2011/entries/tropes/>>.

Bayne, T. and M. Montague, (eds.), 2011, *Cognitive Phenomenology*, New York: Oxford University Press.

Berkeley, G., 1710, "Treatise Concerning the Principles of Human Knowledge", in *The Works of George Berkeley, Bishop of Cloyne*, A.A. Luce and T.E. Jessop (eds.), London: Thomas Nelson and Sons, 9 volumes, 1948–1957.

Bickle, J., 2013, "Multiple Realizability", *The Stanford Encyclopedia of Philosophy (Spring 2013 Edition)*, Edward N. Zalta (ed.), forthcoming URL = <<http://plato.stanford.edu/archives/spr2013/entries/multiple-realizability/>>.

Blackburn, S., 1992, "Filling in Space", *Analysis*, 50(2): 60–65.

Block, N., 1996, "Mental Paint and Mental Latex," in E. Villanueva (ed.), *Philosophical Issues*, 7: Perception: 19-49.

Bird, A. and E. Tobin, 2010, "Natural Kinds", *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2010/entries/natural-kinds/>>.

BonJour, L., 2012, "Epistemological Problems of Perception", *The Stanford Encyclopedia of Philosophy (Winter 2012 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/win2012/entries/perception-episprob/>>.

Bostrom, N., 2003, "Are You Living In a Computer Simulation?", *Philosophical Quarterly*, 53, (211): 243-255.

Bourget, D. and D. Chalmers, 2009, "The PhilPapers Surveys: Preliminary Survey results", n.p., 8 December 2009, web. 20 April 2013.

Brandt, R. B., 1979, *A Theory of the Good and the Right*, Oxford: Oxford University Press.

Brax, D., 2009, *Hedonism as the Explanation of Value*, D. Ph. diss., Lund University.

Broad, C. D., 1925. *The Mind and Its Place in Nature*, London: Routledge & Kegan Paul.

---, 1930, *Five Types of Ethical Theory*, Kegan Paul, Trench, London: Trubner & Co.

Brown, H. I. 1992, "Direct realism, indirect realism, and epistemology", *Philosophy and Phenomenological Research* 52 (2):341-363.

Byrne, A., 2001, "Don't PANIC: Tye's intentionalist theory of consciousness", in *A Field Guide to the Philosophy of Mind* symposium on Tye (2000), available at [http://host.uniroma3.it/progetti/kant/field/tyesymp\\_byrne.htm](http://host.uniroma3.it/progetti/kant/field/tyesymp_byrne.htm)., accessed May 1, 2013.

---, 2006, "Color and the Mind-Body Problem", *Dialectica*, 6 (3):223-244.

Byrne, A. and D. R. Hilbert, 2003, "Color Realism and Color Science", *Behavioral and Brain Sciences*, 26: 1-44.

Byrne, A. and H. Logue (eds.), 2009a, *Disjunctivism: Contemporary Readings*, Cambridge, MA: MIT Press.

---, 2009b "Introduction" in *Disjunctivism: Contemporary Readings*, A. Byrne, and H. Logue, (eds.), Cambridge, MA: MIT Press.

Campbell, J., 2002, *Reference and Consciousness*, Oxford: Oxford University Press.

Casati, R. and J. Dokic, "Sounds," *The Stanford Encyclopedia of Philosophy (Summer 2011 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2011/entries/sounds/>>.

Casati, R. and A. Varzi, 2010, "Events", *The Stanford Encyclopedia of Philosophy (Spring 2010 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2010/entries/events/>>.

Chalmers, D. J., 1996, *The Conscious Mind: In Search of a Fundamental Theory*, New York and Oxford: Oxford University Press.

---, 2005, "The matrix as metaphysics", in *Philosophers Explore the Matrix*, C. Grau (ed.), New York: Oxford University Press.

---, 2010, "Perception and the fall from Eden", in *The Character of Consciousness*, D. Chalmers (ed.), Oxford: Oxford University Press.

*Perceptual Experience*, T. S. Gendler and J. Hawthorne (eds.), Oxford: Oxford University Press.

- , 2004, "The representational character of experience", *The Future for Philosophy*, In B. Leiter (ed.), Oxford: Oxford University Press.
- , 2009, "The Two-Dimensional Argument Against Materialism", in *Oxford Handbook to the Philosophy of Mind*, B. P. McLaughlin & S. Walter (eds.), Oxford: Oxford University Press.
- Chisholm, R.M., 1957, *Perceiving: A Philosophical Study*, Ithaca: Cornell University Press.
- , 1970, 'Events and Propositions', *Noûs*, 4, 15–24.
- Clark, A., 2005, "The twisted matrix: Dream, simulation, or hybrid?", in *Philosophical Essays on the Matrix*, C. Grau (ed.). New York: Oxford University Press.
- Cornman, J., 1971, *Materialism and Sensations*, New Haven, Conn.: Yale University Press.
- , 1975, *Perception, Common Sense, and Science*, New Haven, Conn.: Yale University Press.
- Crane, T., 1998, "Intentionality" in *Routledge Encyclopedia of Philosophy*, E.J. Craig (ed.), London: Routledge.
- , 2000, "The Origins of Qualia", in *History of the Mind-Body Problem*, T. Crane and S. Patterson (eds.), London: Routledge.
- , 2011, "The Problem of Perception", *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/perception-problem/>.
- Crisp, R., 2006, *Reasons and the Good*, Oxford: Clarendon Press.
- Dennett, D., 1991. *Consciousness Explained*, New York: Little, Brown.
- , 1995, 'The Unimagined Preposterousness of Zombies', *Journal of Consciousness Studies*, 2: 322–6.

- , 1999, "Quining Qualia" in *The Nature of Consciousness - Philosophical Debates*, Block, Flanagan and Güzeldere (eds.), Cambridge, MA: MIT press, 619-620.
- Deutsch, M., 2005, "Intentionalism and intransitivity", *Synthese* 144:1-22.
- Drake, S., 1957, *Discoveries and Opinions of Galileo*, Garden City, NY: Doubleday.
- Dretske, F., 1995, *Naturalizing the Mind*, Cambridge, MA: Bradford Books / MIT Press.
- , 1993, "Conscious experience", *Mind* 102 (406):263-283.
- Ducasse, C.J., 1942, "Moore's Refutation of Idealism" in P. Schilpp (ed.) *The Philosophy of G.E. Moore*, 223-52, Chicago: Northwestern University Press.
- Dummett, M., 1975, "Wang's paradox", *Synthese* 30: 301-24.
- Evans, G., 1978, "Can there be Vague Objects?", *Analysis*, 38: 208.
- Everett, A., 1996, "Qualia and Vagueness", *Synthese* 106: 205-226.
- Feldman, F., 1997, "On the Intrinsic Value of Pleasures", *Ethics*, 107:448-466.
- Flohr, H., 2000, "NMDA-receptor-mediated computational processes and phenomenal consciousness", in Thomas Metzinger (ed.), *Neural Correlates of Consciousness*, Cambridge MA: MIT Press.
- Fish, W. C., 2004, "The direct/indirect distinction in contemporary philosophy of perception", *Essays in Philosophy* 5 (1):1-13.
- , 2008, "Disjunctivism, Indistinguishability, and the Nature of Hallucination", in *Disjunctivism: Perception, Action, Knowledge*, F. Macpherson and A. Haddock (eds.), Oxford: Oxford University Press, 144-167.
- , 2009, *Perception, Hallucination, and Illusion*, Oxford: Oxford University Press.
- , 2010, *Philosophy of Perception: A Contemporary Introduction*, Routledge.

- Frege, G., 1966, "Function and Concept," translated by P. T. Geach, in *Translations from the Philosophical Writings of Gottlob Frege*, P. Geach and M. Black (eds.), Oxford: Blackwell, 21–41.
- Fodor, J., 1990, "Making Mind Matter More," in Fodor, J., *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press.
- Forrest, P., 2005, "Universals as Sense-data", *Philosophy and Phenomenological Research*, 71 (3): 622 - 631.
- Goldman, A., 1986, *Epistemology and Cognition*, Cambridge, Mass.: Harvard University Press.
- Goodman, N., 1951, *The Structure of Appearance*, Cambridge, MA: Harvard University Press.
- Gosling, J.C.B., 1969, *Pleasure and Desire: The Case for Hedonism Reviewed*, Oxford: Oxford University Press.
- Graff, D., 2001, "Phenomenal Continua and the Sorites", *Mind* 110: 905–935.
- Grice, H.P., 1961, "The Causal Theory of Perception" *Proceedings of the Aristotelian Society Supplementary Volume 35*: 121-52.
- Van Gulick, R., 2005, "So many ways of saying no to Mary", in Ludlow et al (2005).
- Haddock, A. and F. Macpherson, 2008, *Disjunctivism: Perception, Action, Knowledge*, Oxford: Oxford University Press.
- Hall, R. J., 1989,. "Are Pains Necessarily Unpleasant?", *Philosophy and Phenomenological Research*, 49(4): 643–659.
- Hameroff, S. R., 1998, "Fundamental geometry: The Penrose-Hameroff "Orch OR" model of consciousness", in *The Geometric Universe - Science, geometry and the work of Roger Penrose*, N. S. A. Huggett and L. J. Mason, K. P. Tod, S. T. Tsou, and N. M. J. Woodhouse (eds.), 135-160.
- Harman, G., 1990, "The Intrinsic Quality of Experience" in *Philosophical Perspectives*, J. Tomberlin (ed.) 31–52, Atascadero: Ridgeview.

Hilbert, D., 2004, "Hallucination, sense-data, and direct realism", *Philosophical studies*, 120: 185-191.

Hill, C. S., 1997, 'Imaginability, Conceivability, Possibility and the Mind-Body Problem', *Philosophical Studies*, 87: 61-85.

Hill, C. S., and B. P. McLaughlin, 1999, 'There are Fewer Things in Reality Than Are Dreamt of in Chalmers's Philosophy', *Philosophy and Phenomenological Research*, 59: 446-454.

Hinton, J.M., 1973, *Experiences: An Inquiry into Some Ambiguities*, Oxford: Clarendon Press.

Huemer, M., 2001, *Skepticism and the Veil of Perception*, Lanham, Md.: Rowman & Littlefield.

---, 2011, "Sense-Data", *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/sense-data/>.

Hume, D., 1758, "An Enquiry Concerning Human Understanding", in *Enquiries concerning Human Understanding and concerning the Principles of Morals*, L. A. Selby-Bigge (ed.), 3rd edition revised by P. H. Nidditch, Oxford: Clarendon Press, 1975.

---, Hume, D., 2000, *A Treatise of Human Nature*, Oxford: Oxford University Press.

Huxley, T.H., 1999, "On the hypothesis that animals are automata, and its history," excerpted in *The mind-body problem - An opinionated introduction*, D. M. Armstrong, 1999, Oxford: Westview press.

Jackson, F., 1977, *Perception: A Representative Theory*, Cambridge: Cambridge University Press.

---, 1982, "Epiphenomenal qualia", *Philosophical Quarterly*, 32:127-136.

---, 1986, "What Mary Didn't Know", *Journal of Philosophy*, 83: 291-5.

---, 1998, "Postscript on Qualia," in his *Mind, Method and Conditionals: Selected Essays*, London: Routledge, 76-9.

Jackson, F. and R.J. Pinkerton, 1973, "On an Argument Against Sensory Items", *Mind* 82: 269-272.

Johnston, M., 2004, "The Obscure Object of Hallucination", *Philosophical Studies*, 103: 113-83.

Kirk, R., "Zombies", *The Stanford Encyclopedia of Philosophy (Summer 2012 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2012/entries/zombies/>>.

Levine, J., 1983, "Materialism and Qualia: the Explanatory Gap.", *Pacific Philosophical Quarterly* 64: 354-361.

Locke, J., 1689, *An Essay Concerning Human Understanding*, P. Nidditch (ed.), Oxford: Clarendon: 1975.

Kagan, S., 1998, "Rethinking Intrinsic Value", *The Journal of Ethics*, 2: 277-97.

Kant, I., 2003, *Critique of Pure Reason*, translated by N. Kemp Smith, with a new preface by H. Caygill, New York: Palgrave Macmillan.

Kim, J., 2006, *Philosophy of Mind (Second Edition)*, Boulder: Westview Press.

---, 2005, *Physicalism, or Something Near Enough*, Princeton: Princeton University Press.

Knobe, J., K. D. Olum, and A. Vilenkin, 2006, "Philosophical implications of inflationary cosmology", *British Journal for the Philosophy of Science* 57 (1):47-67.

Kripke, S., 1972/80, *Naming and Necessity*, Oxford: Blackwell. (Revised and enlarged version of 'Naming and Necessity', in *Semantics of Natural Language*, D. Davidson and G. Harman (eds.), Dordrecht: D. Reidel, 253-355.)

Langsam, H., 1997, "The Theory of Appearing Defended", *Philosophical Studies*, 120: 33-59.

Levin, Janet, 2010, "Functionalism", *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2010/entries/functionalism/>>.

Lewis, D., 1983, "New Work for a Theory of Universals", *Australasian Journal of Philosophy*, 61: 343–77.

---, 1988a, "Veridical Hallucination and Prosthetic Vision", in *Perceptual Knowledge*, J. Dancy (ed.), Oxford: Oxford University Press.

---, 1988b, "What Experience Teaches", *Proceedings of the Russellian Society*, 13: 29–57.

Loar, B., 1990/1997, 'Phenomenal States', in *Philosophical Perspectives*, Vol. 4, J. Tomberlin (ed.), Atascadero: Ridgeview.

---, Loar, B., 2003, "Transparent experience and the availability of qualia", in *Consciousness: New Philosophical Perspectives*, Q. Smith and A. Jokic (eds.), Oxford University Press.

Lowe, E. J., 2002, *A Survey of Metaphysics*, Oxford: Oxford University Press.

Ludlow, P., Y. Nagasawa, and D. Stoljar, (eds.), 2005, *There is something about Mary: essays on phenomenal consciousness and Frank Jackson's knowledge argument*, Cambridge, MA: MIT Press.

Lycan, W. G., 1987, *Consciousness*, Cambridge, MA: Bradford Books / MIT Press.

---, 2001, "The case for phenomenal externalism", *Philosophical Perspectives* 15 (15):17-35.

---, 2008, "Representational Theories of Consciousness," *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.),  
URL =  
<<http://plato.stanford.edu/archives/fall2008/entries/consciousness-representational/>>.

Mackie, J., 1977, *Ethics: Inventing Right and Wrong*, New York: Penguin.

Martin, M.G.F., 2002, "The Transparency of Experience", *Mind and Language*, 17: 376–425.

---, 2004, "The Limits of Self-Awareness", *Philosophical Studies*, 120: 37–89.

---, 2006, "On Being Alienated", in *Perceptual Experience*, T. S. Gendler and J. Hawthorne (eds.), Oxford: Oxford University Press, 354–410.

Maund, B., 2012, "Color," *The Stanford Encyclopedia of Philosophy (Winter 2012 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/win2012/entries/color/>>.

Maurin, A-S., 2002, *If Tropes*, Dordrecht: Kluwer Academic Publishers.

McDowell, J., 1982, "Criteria, Defeasibility and Knowledge", *Proceedings of the British Academy*, 68: 455–79.

Milner, D. and M. Goodale, 1995, *The Visual Brain in Action*, Oxford: Oxford University Press.

Moore, G. E., 1903, "The refutation of idealism", *Mind*, 12 (48):433-453.

---, 1953. *Some Main Problems of Philosophy*, London: George, Allen and Unwin.

Nida-Rümelin, Martine, 2010, "Qualia: The Knowledge Argument", *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2010/entries/qualia-knowledge/>>.

Nozick, Robert, 2005, "The experience machine", excerpted in *Philosophers Explore the Matrix*, C. Grau (ed.), New York: Oxford University Press.

O'Brien, D., 2013, "Objects of Perception", *The Internet Encyclopedia of Philosophy*, ISSN 2161-0002, <http://www.iep.utm.edu/>, 30 April 2013.

O'Neill, J., 1992, "The Varieties of Intrinsic Value", *The Monist*, 75(2): 119-37.

Palmer, S. E., 1999, *Vision science: Photons to phenomenology*, Cambridge, MA: MIT Press.

Papineau, D., 2002, *Thinking about Consciousness*, Oxford: Clarendon Press.

Parfit, D., 1984, *Reasons and Persons*, Oxford: Clarendon Press.

- Pautz, A., ms, "Color eliminativism", available at: <https://webspace.utexas.edu/arp424/www/elim.pdf>, accessed May 1, 2013.
- Pitcher, G., 1971, *A Theory of Perception*. Princeton, NJ: Princeton University Press.
- Ploner, M., H. J. Freund and A. Schnitzler, 1999, "Pain Affect without Pain Sensation in a Patient with a Postcentral Lesion," *Pain*, 81(1/2): 211–214.
- Price, H. H., 1950, *Perception, 2nd edition*, London: Methuen.
- Price, R., 2006, *The Way Things Look*, D. Ph. diss., Oxford University.
- Quine, W. V. O., "On What There Is," in *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 1–19.
- , 1960, *Word and Object*, Cambridge (MA): MIT Press.
- Rabinowicz, W. and T. Rønnow-Rasmussen, 1999, "A Distinction in Value: Intrinsic and For Its Own Sake", *Proceedings of the Aristotelian Society*, 100: 33-49.
- Rachels, S. 2000, "Is unpleasantness intrinsic to unpleasant experiences?", *Philosophical Studies* 99 (2):187-210.
- Rapaport, W. J., 1978, "Meinongian Theories and a Russellian Paradox," *Noûs*, 12: 153–180.
- Reicher, M., 2012, "Nonexistent Objects," *The Stanford Encyclopedia of Philosophy (Winter 2012 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/win2012/entries/nonexistent-objects/>>.
- Revonsuo, A., 2000, *Inner Presence: Consciousness As a Biological Phenomenon*, Cambridge, MA: MIT Press.
- Rey, G., 1983, "A Reason for Doubting the Existence of Consciousness," in *Consciousness and Self-Regulation Vol 3*, R. Davidson, G. Schwartz and D. Shapiro (eds.), New York, Plenum: 1-39.
- , 1988, "A Question About Consciousness," in *Perspectives on Mind*, H. Otto & J. Tuedio (eds.), Dordrecht: Reidel, 5–24.

Robinson, H., 1985, "The General Form of the Argument for Berkeleyian Idealism", in *Essays on Berkeley: A Tercentennial Celebration*, J. Foster and H. Robinson (eds.), Oxford: Clarendon Press, 163–186.

---, 1994, *Perception*, London: Routledge.

Robinson, W., 2012, "Epiphenomenalism", *The Stanford Encyclopedia of Philosophy (Summer 2012 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2012/entries/epiphenomenalism/>>.

Rodriguez-Pereyra, G., 2011, "Nominalism in Metaphysics", *The Stanford Encyclopedia of Philosophy (Fall 2011 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2011/entries/nominalism-metaphysics/>>.

Rosen, G., 2012, "Abstract Objects", *The Stanford Encyclopedia of Philosophy (Spring 2012 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2012/entries/abstract-objects/>>.

Russell, B., 1914, "The Relation of Sense-Data to Physics", *Scientia*, 1–27.

---, 1927a, *The Analysis of Matter*. Reprinted: New York: Dover, 1954.

---, 1927b, *An Outline of Philosophy*, Reprinted: London: Allen and Unwin, 1979.

---, 1959, *My Philosophical Development*, Reprinted: New York: Simon and Schuster 1975.

---, 1987, *The Problems of Philosophy*, Oxford: Oxford University Press.

Samuelsson, L., 2008, *The Moral Status of Nature: Reasons to Care for the Natural World*, D. Ph. diss., Umeå University.

Schwitzgebel, E., 2007, "Do you have constant tactile experience of your feet in your shoes? Or is experience limited to what's in attention?", *Journal of Consciousness Studies*, 14(3): 5–35.

Shoemaker, S. 1996a, *The first-person perspective and other essays*, Cambridge: Cambridge University Press.

---, 1996b, "Self-knowledge and 'inner sense': Lecture I", in Shoemaker 1996, 201–223.

- Siegel, S., 2006, "Which Properties are Represented in Perception?" in *Perceptual Experience*, T. S. Gendler and J. Hawthorne (eds.), Oxford: Oxford University Press, 481–503.
- , 2011, "The Contents of Perception," *The Stanford Encyclopedia of Philosophy (Winter 2011 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/win2011/entries/perception-contents/>>.
- Siewert, C., 1998, *The Significance of Consciousness*, Princeton University Press.
- Singer, W., 1999, "Neuronal synchrony: a versatile code for the definition of relations", *Neuron*, 24: 49-65.
- Smart, J.J.C., 1959, 'Sensations and Brain Processes', *Philosophical Review*, 68: 141–156.
- Smith, A. D., 2006,. "In defence of direct realism", *Philosophy and Phenomenological Research* 73 (2):411-424.
- Smythies, J., 2003, "Space, time and consciousness", *Journal of Consciousness Studies* 10 (3):47-56.
- Snowdon, P.F., 1980–81, "Perception, Vision and Causation", *Proceedings of the Aristotelian Society, New Series*, 81: 175–92.
- Soteriou, M., 2010, "The Disjunctive Theory of Perception," *The Stanford Encyclopedia of Philosophy (Winter 2010 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/win2010/entries/perception-disjunctive/>>.
- Stoljar, D., 2001a, "Two Conceptions of the Physical", *Philosophy and Phenomenological Research*, 62: 253–281.
- , 2001b, "The Conceivability Argument and Two Conceptions of the Physical", *Philosophical Perspectives*, 15: 393–413.
- , 2004, "The Argument from Diaphanousness" in *New Essays in the Philosophy of Language and Mind: Canadian Journal of Philosophy Supplementary Volume*, M. Ezcurdia, R. Stainton and C. Viger (eds.) 341–90, Calgary: University of Calgary Press.

- , 2006, *Ignorance and Imagination. The Epistemic Origin of the Problem of Consciousness*, Oxford: Oxford University Press.
- , 2009, "Physicalism", *The Stanford Encyclopedia of Philosophy (Fall 2009 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2009/entries/physicalism/>>.
- Stubenberg, L., 2010, "Neutral Monism", *The Stanford Encyclopedia of Philosophy (Spring 2010 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2010/entries/neutral-monism/>>.
- Sturgeon, S., 2008, "Disjunctivism About Visual Experience", in *Disjunctivism: Perception, Action, Knowledge*, F. Macpherson and A. Haddock (eds.), Oxford: Oxford University Press, 112–143.
- Tegmark, M., 2003, "Parallel Universes", *Scientific American*, May, 41–51.
- Thompson, B. J., 2008, "Representationalism and the argument from hallucination". *Pacific Philosophical Quarterly* 89 (3): 384-412.
- Travis, C., 2004. "The Silence of the Senses", *Mind*, 113 (449): 57–94.
- Tye, M., 1995. *Ten Problems of Consciousness*. Cambridge, MA: Bradford Books/MIT Press.
- , 1997, "A Representational Theory of Pains and their Phenomenal Character", in *The Nature of Consciousness: Philosophical Debates*, N. Block, O. Flanagan and G. Güzeldere, (Eds.), Cambridge, MA: MIT Press.
- , 2000, *Consciousness, Color, and Content*, Cambridge, Mass: The MIT Press.
- , 2005, "In Defense of Representationalism: Reply to Commentaries", in *Pain: New Essays on its Nature and the Methodology of its Study*, M. Aydede (ed.) Cambridge, MA: MIT Press, 2005,163-75.
- , 2006, "Another Look at Representationalism about Pain", in *Pain: New Essays on Its Nature and the Methodology of Its Study*, M. Aydede (Ed.), Cambridge MA: MIT Press.

---, 2009, "Qualia", *The Stanford Encyclopedia of Philosophy (Summer 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2009/entries/qualia/>.

Ward, A., 1976, "Direct and indirect realism", *American Philosophical Quarterly* 13 (October): 287-294.

Wilkes, K., 1988, "Yishi, Duh, Um and Consciousness," in *Consciousness in Contemporary Science*, A. Marcel, and E. Bisiach (eds.), Oxford: Oxford University Press.

---, 1995, "Losing Consciousness," in Metzinger, T. (ed.), *Consciousness and Experience*, Ferdinand Schoningh.

Williamson, T., 2000. *Knowledge and its Limits*, Oxford: Oxford University Press.

Yaffe, G. and Nichols, R, 2009, "Thomas Reid", *The Stanford Encyclopedia of Philosophy (Winter 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2009/entries/reid/>.

Zeki, S., 1983, "Colour coding in the cerebral cortex: The reaction of cells in monkey visual cortex to wavelengths and colours", *Neuroscience*, 9:741-65.



## UMEÅ STUDIES IN PHILOSOPHY

Department of Historical, Philosophical and Religious Studies  
Umeå University, Sweden

The series UMEÅ STUDIES IN PHILOSOPHY consists mainly of research communications (monographs, dissertations and collections of essays) in philosophy from the Department of Historical, Philosophical and Religious Studies at Umeå University. The series is not distributed through the bookstores, but individual volumes can be obtained from the department. Orders should be sent to Umeå Studies in Philosophy, Department of Historical, Philosophical and Religious Studies, Umeå University, SE-901 87 Umeå, Sweden.

1. JONAS NILSSON: *Rationality in Inquiry: On the Revisability of Cognitive Standards*, 2000.
2. PER NILSSON: *Naturen, vetenskapen & förnuftet: Upplysningens dialektik och det andra moderna*, 2001.
3. JAYNE M. WATERWORTH: *Living in the Light of Hope: An Investigation into Agency and Meaning*, 2001.
4. ANDERS ODENSTEDT: *Cognition and Cultural Context: An Inquiry Into Gadamer's Theory of Context-Dependence*, 2001.
5. RÖGNVALDUR INGTHORSSON: *Time, Persistence, and Causality: Towards a Dynamic View of Temporal Reality*, 2002.
6. BENGT LILIEQUIST: *Ludwik Flecks jämförande kunskapsteori*, 2003.
7. PETER NILSSON: *Empathy and Emotions: On the Notion of Empathy as Emotional Sharing*, 2003.
8. ANDERS BERGLUND: *From Conceivability to Possibility: An Essay in Modal Epistemology*, 2005.
9. LARS SAMUELSSON: *The Moral Status of Nature: Reasons to Care for the Natural World*, 2008.
10. EBBA GULLBERG: *Objects and Objectivity: Alternatives to Mathematical Realism*, 2011.

11. JESPER ÖSTMAN: *It's All in the Brain: A Theory of the Qualities of Perception* 2013.