The explanatory gap problem: how neuroscience might contribute to its solution

D i s s e r t a t i o n

zur Erlangung des akademischen Grades

Doctor philosophiae (Dr. phil.)

Eingereicht an der Philosophischen Fakultät I

der Humboldt-Universität zu Berlin

von Daniel Kostić , Dipl.-Phil.

Der Präsident der Humboldt-Universität zu Berlin Prof. Dr. Dr. h.c. Christoph Markschies

Der Dekan der Philosophischen Fakultät I Prof. Dr. Michael Seadle

Gutachter:

1. Prof. Dr. Michael Pauen

2. Prof. Dr. David Papineau

Datum der Promotion: 12.07.2011

© 2010 Daniel Kostić

Dedicated to the memory of the late professor Nikola Grahek, who sparked my interest in philosophy of mind

Table of Contents

Abstract

This thesis evaluates several powerful arguments that not only deny that brain states and conscious states are one and the same thing, but also claim that such an identity is unintelligible. I argue that these accounts do not undermine physicalism because they don't provide any direct or independent justification for their tacit assumptions about a link between modes of presentation and explanation. In my view intelligibility of psychophysical identity should not be based exclusively on the analysis of meaning. The main concern then should be why expect that fully intelligible explanation must be based on the descriptions of the causal roles as modes of presentation. To this effect I propose that we examine "psychological concepts". The psychological concepts are concepts that use descriptions of the functional roles but are about qualities of our experiences. I propose to analyze them in quality space models in order to unveil why phenomenal concepts are expected to refer via descriptions of the causal or functional roles.

The quality space should be understood here as a multidimensional space consisting of several axes of relative similarity and differences among the structures of ordering in different modalities of conscious experience. On my proposal it is possible that some axes in the quality space consist of their own quality spaces so we could "zoom in" and "zoom out" into the descriptions of the functional roles and see more clearly what the explanation of certain aspects of consciousness looks like when thought of in terms of psychological concepts.

Acknowledgements

Completing a PhD dissertation is by no means an easy endeavour. It is a journey one cannot embark on and successfully finish without help and support from others. I am truly grateful for the opportunity to meet all the people who in one way or another helped me bring this journey to an end. I would thus like to thank to all of my friends and colleagues for their generous support and encouragement during my PhD.

Without financial support from the Villigst Foundation I could not have carried this project through. I am very proud to have been awarded their prestigious and tremendously competitive doctoral scholarship. Their generosity, both in terms of funding and in terms of organizational support and genuine care, are tokens of the privilege I have been given as a fellow. I would like to express my special gratitude to Mrs. Karin Lühmann from the Villigst Foundation. Her kindness, confidence and experience made the communication and administration of my scholarship seamless.

I am grateful to Michael Pauen who is my main PhD advisor, without whom I would not have been able to start this project.

I am especially indebted to Professor David Papineau for his generosity with time and advice. His work in general and my tutorials with him at King's College London were tremendously important for my philosophical development.

I would also like to thank Michael Schütte for his kindness and for commenting many papers that I wrote in the course of my PhD.

I would like to thank Paula Droege for her thorough and helpful comments on my key arguments and for the encouragement.

I am thankful to all members of Professor Pauen's colloquium at the Humboldt University in Berlin for their helpful comments on various papers that I presented at Berlin's school of Mind and Brain. I am particularly grateful to Daniel Friedrich. I also wish to thank to all members of the Mind and Metaphysics Study groups at Kings College London as well as to

all members of the Advanced Research Seminar at Kings College London for very stimulating discussions, from which I learned a great deal.

My special gratitude goes to Robert Templing for his thorough and insightful comments on my key arguments. His attention to details and kindness are astonishing. I am indebted to Manolo Martinez for discussing arguments in my dissertation and providing extremely helpful and insightful comments on various points in it. I thank Tyron Goldschmidt, Jennifer Wright and Luke Brunning for kindly agreeing to proofread my dissertation. Their meticulous work on proofreading and their comments and encouragement are tremendously important to me.

I am thankful to Benjamin Waters, Silvia Cason, Marc Borner and Stephen Tiley for a wonderful friendship, for their interest in my work, very interesting discussions and encouragement. I am particularly grateful to Ben who read, commented and proofread many of the papers and materials that I wrote in the course of my PhD.

I am thankful to my undergraduate mentor Professor Živan Lazović for everything he has done for me and for all the advice and encouragement he gave me.

Finally, I am grateful to Slavica Radisić for her generosity, support, encouragement and patience. I would also like to thank Slavica's family for the support and understanding throughout my studies and especially during the last few months of my PhD. Last but not least, I would like to thank my family for their support and for letting me know how proud they feel about me.

If I omitted to mention someone I beg for forgiveness, for it was not my intention.

Introduction

In this work I shall be concerned with questions of the following sort: why are physical states of the brain always accompanied by some conscious states, or why there is something it is like to undergo certain experiences or states at all. These questions capture the essence of the famous "explanatory gap". It is explicitly claimed, in the premises of some of the classical arguments in philosophy of mind, that we will never be in a position to understand how and why "brain states give rise to phenomenal states" even if psychophysical identity were true (Levine 1983, 1993, 2001). Although they initially sound simple these questions are by no means easy. In fact, David Chalmers calls these issues the hard problem of consciousness (Chalmers 1995). What makes this problem so hard is the fact that two distinct kinds of access to consciousness seem incompatible, one from the first person perspective and the other from the third person perspective. That is to say, we can have knowledge about our experiences from the third person perspective but that would not allow us to know what it is like to undergo that experience, and *vice versa*. Furthermore, to some philosophers, like David Chalmers (Chalmers 1996) or Frank Jackson (Jackson 1982) this incompatibility reflects a much deeper distinction in nature. They maintain that two epistemic perspectives of consciousness are anchored in two metaphysically distinct kinds of properties.

What does it actually mean to claim that the different ways in which we think of consciousness seem incompatible?

On the one hand we can think about consciousness from the third person perspective. This would include all the knowledge we have from neuroscience, biology, physics, medicine, cognitive science and the like; about the workings of the brain and its environment. On the other hand, we can think about consciousness from the first person perspective. This way of thinking about consciousness can best be described as that it looks somehow to be in such a state (Nagel 1974). This situation is peculiar to consciousness and it is commonly referred to as the *epistemic gap*. This gap poses a problem to some philosophers (Jackson 1982; Chalmers 1996, 2000, 2009) because they believe that what we

are thinking of from the first person perspective and from the third person perspective cannot be the same thing, or at least it cannot be intelligible how they could be the same thing (Levine 1983, 1993, 2001). They argue that, actually, the epistemic gap reflects a much deeper gap in nature and that is an *ontological gap.* Or, at the very least it is not intelligible as how these two distinct perspectives are about one and the same thing; such a position is known as *the explanatory gap* (Levine 1983).

In arguing against the ontological gap, some philosophers, such as Joseph Levine (Levine 1983, 1993, 2001), argue that drawing metaphysical conclusions from epistemic premises is not justified in this case and that we cannot infer the ontological gap from the mere existence of the epistemic gap. This means that the existence of the epistemic gap by no means leads to the metaphysical conclusion that these two ways of thinking about consciousness reflect two metaphysically distinct kinds of properties. However, he claims, this should not put at ease the philosophers who maintain that these two ways of thinking about consciousness are basically about one and the same thing, because even if it were true that it is about the same thing, we can't really *understand* how this could be. This is the explanatory gap problem (Levine 1983).

In this work, among other things, I urge to make a sharper distinction between *the epistemic gap* and *the explanatory gap*. The mere existence of the epistemic gap is not a reason to suppose that it is not explicable. Perhaps we should not expect to have the same pattern of explanation of the epistemic gap in a way some philosophers such as Levine, Chalmers and Jackson would argue we should but, rather, that an explanation might come from a source other than the analysis of meaning. I argue in chapters four and five that the intelligibility of explanation of the psychophysical identity follows a constraint of the vagueness of phenomenal concepts and in principle cannot come directly from the identity alone. Apart from that, proper identities are not in fact explanatory (Block and Stalnaker 1999; Papineau 2002). Their purpose is to transfer explanatory and causal force onto some further properties (Block and Stalnaker 1999). David Papineau (Papineau 2002) argues that psychophysical identity[1] is non-explanatory, as all proper identities are, but that this does

[1] The claim that mental states and brain states are one and the same thing. I will come to this issue in short while below, when I will be discussing the causal argument for physicalism.

not mean that the explanatory gap cannot be explained. This position will be the main topic of the third and the fourth chapters, so I will not pursue it right now. What is really important to emphasize here is that the sharp distinction between the epistemic gap and the explanatory gap is very important because it minimizes confusion about the intelligibility of explaining psychophysical identity. This distinction allows for claiming that there is an epistemic gap but no explanatory gap.

These are all very complicated issues and I will be discussing them in much greater detail in chapters four and five.

How to formulate physicalism

However, before we get into considering the arguments for these claims we need to ask ourselves whether there is any reason to believe that these two ways of thinking about consciousness could be about the same thing? I think there is a very good reason to maintain that they are about the same thing. I refer here to Papineau's causal argument for physicalism (Papineau 2002). I think we need not examine each individual variety of physicalism and then pick the most suitable candidate. To do such a thing we would need to set the criteria for what we are looking for etc., which sounds circular and unpromising on the face of it[2]. A better way is to see what the most basic assumptions about consciousness that we are allowed to make are and then just to draw the consequences from them. I propose we do exactly that. Papineau's causal argument for physicalism goes like this:

"1. Conscious mental occurrences have physical effects.

2. All physical effects are fully caused by purely *physical* histories.

3. The physical effects of conscious causes are not always overdetermined by distinct causes.

4. Therefore, physicalism is true" (Papineau 2002, pp. 17-18).

[2] It is perhaps circular because we start with the question "what is physicalism" and we try to answer it by asking what are the criteria for something to be physicalism. To know such criteria we need to know what phyiscalism is beforehand.

This argument brilliantly and elegantly encapsulates the very idea of physicalism. At first pass premises 1. and 2. alone seem to suggest that some physical effects, say our behaviour, could have two distinct causes. Premise 1. acknowledges the causal efficacy of qualia. Premise 2. is what Papineau calls the "completeness of physics" thesis " (Papineau 2002, p. 17). The purpose of "the completeness of physics" thesis is to assure that all causes are physical causes. Finally, the premise 3. expresses the fact that certain effects are not caused by distinct causes, namely: the physical causes and the conscious causes. In other words, the premise 3. prevents overdetermination. Although overdetermination does not sound so odd in other domains, in the case of consciousness it seems very strange. Consider this example: a man dies as a consequence of being struck by lightning and being shot by a gun bullet simultaneously. It is perfectly intelligible to think that his death could be an effect of two causes: lightning and a lethal bullet wound, or at least that these two causes might have been a cause of his death individually/independently. In other words, he could have died even if one of the causes was absent. However, in the case of mental causation things do not seem so straightforward. It does sound odd to say, for example, that my pain *qua* mental state caused me to move my hand and that my pain *qua* brain state caused me to move my hand simultaneously. Or, to make it even more queer, consider this: I moved my hand because I felt pain *qua* mental state but I would not have moved it if it had been pain *qua* brain state, or I moved my hand because I felt pain *qua* brain state but I would not have moved it if it had been pain *qua* mental state. To prevent such odd situations we need a premise that would constrain overdetermination.

If conscious and physical causes are not distinct, as the premise 3. claims, then the only remaining possibility is to conclude that conscious and physical causes are identical. In this way we directly get the most plausible characterization of physicalism, in which a discussion of other varieties of physicalism plays no significant role.

The causal argument can be rephrased so it could generate the same strong conclusion even on very different accounts of causation. However, by modifying some of the premises accordingly we get the same conclusion even on very different account of causation.

What is more important is the question: what do we get when some or all premises of the argument are denied? So let's briefly walk through some attempts to refute each one of them and see whether these attempts bear any plausibility.

Denying premise 1. amounts to saying that conscious states don't have physical effects. A classic example of this position is epiphenomenalism. On this view conscious causes exist alongside physical causes but have no effect on them. Perhaps they can have some effects on other conscious states, but that does not change much with regards to the main point of this approach. Jackson (Jackson 1982) argued that if conscious states exist, and have no causal effects on the brain, then consciousness is a by-product of these brain states, i.e. consciousness is epiphenomenal. However, if physical world is causally closed[3], then a very important question arises: can science afford to keep these "causal danglers"? I will be discussing epiphenomenalism in the light of an inconsistency objection in the third chapter. It suffices to say, for the time being, that it does not seem coherent to hold both that epiphenomenalism is true and that mental causes are physically inefficacious, because in that case we would not be able to detect them. Although epiphenomenalism seems to be a conceptually coherent position, it cannot accommodate any kind of empirical testing in principle. In addition to that epiphenomenalism leaves causal danglers which in effect renders the position implausible.

Premise 2., the completeness of physics thesis, seems harder to challenge as it depends on our understanding of the notion "physics". Complete physics can't be something that is being currently studied at universities and in school textbooks nor can it be a future physics in its ideal state. The former is obviously wrong because current physics is by no means complete. The latter is also wrong understanding of "complete physics" because we don't know yet whether it will be complete ever and how this completeness should look like. So what is it then? Papineau (Papineau 2002, p. 41) argues that instead of asking what a complete physics would include we should ask what it wouldn't include. That is a more important question as we might never be in a position to know what a complete physics will consist of. Instead, we could define the completeness of physics in terms which have nothing

[3] The completeness of physics thesis.

to do with mentally identifiable entities. The mental, in this sense, can be understood as a realm that can be identified via some distinct way of picking out its properties[4]. So we can say that the physical is everything that is not mentally-identifiable. In this way, we may not know everything about physics but still be able to generate the same conclusion of the causal argument.

Papineau (Papineau 2002, p. 41) argues further that the notion "physics" should also be something non-biologically identifiable. In this way we get a stronger premise which can be thought of as rendering the realm of physical as something that might be called "inanimate" and, again, generate the same conclusion that mental and physical causes are one and the same thing.

Denying the premise 3. is a tricky business because it would imply that two distinct causes simultaneously could have the same effect. That is to say, physical effects of conscious causes are always overdetermined. Some philosophers argue that such a setting is not implausible because perhaps that is the way nature or evolution make it sure that the effect really takes place. It is like belt and braces principle (Mellor 1995). But this position seems to generate more problems than solutions. For example, now we have to explain why are particular distinct causes always accompanied by each other or why would particular effect require particular combination of distinct causes, which brings us back at where we started. In this sense accepting overdetermination does not seem to lead anywhere.

After having considered possible refutations of all three premises we are now in a position to state the causal argument as follows:

"5. Conscious causes have inanimate effects.

6. Inanimate effects always have full inanimate causes.

7. So conscious properties must be identical with (or realized by) inanimate properties" (Papineau 2002, p. 42).

---

[4] Details of such a way should not concern us here, because the argument will generate the same conclusion nevertheless.

In this way, as I said earlier, we get the same conclusion with both weaker and stronger definitions of the notion "physics". The really important point, in relation to the conclusion that mental and physical causes are identical, is to make sure that premise 1. says that mental causes have inanimate effects. If physical/inanimate effects can only have physical/inanimate causal histories, then we have no choice but to conclude that mental causes and physical/inanimate causes are the same thing. We can say that on this account physicalism can be formulated as the identity thesis according to which mental properties are identical with physical properties or some higher order physically realized properties.

I am very much inclined to think that this characterisation of physicalism is the most plausible one and that the causal argument is really hard to refute. We shall see in the first chapter how different views about what physicalism should be committed to determine how do we formulate it as a general ontological thesis. And I will also be defending this version of physicalism throughout my dissertation.

Indeed, in principle, we have no a priori or ideological reasons to be physicalists, and perhaps some other doctrines are better suited for explaining the epistemic gap. However, when confronted with the causal argument we have no choice but to accept physicalism as an identity thesis. All we need to do is to isolate the most basic and fundamental assumptions about consciousness and then the right characterization shows itself. This is the main reason why I did not include any preliminary considerations about varieties of physicalism in the introduction.

Having settled the issue of how to formulate physicalism and how it is to be understood and defended I think we can now proceed with laying out the structure of this dissertation. Before doing so, I need to say something about my general approach in the thesis.

Overall approach

As the title of my dissertation suggests, I should be mainly concerned with the question as to how neuroscience might contribute to the solution of the explanatory gap problem. However, I discuss that particular issue right at the very end. I decided to apply such a strategy because these issues are very complicated and I wanted to build up and

branch out towards my key argument very carefully and gradually. I believe such an approach will also help present the ideas in the most systematic way.

In my dissertation I have decided to follow a specific strategy of laying out ideas and positions. I start with some of the classical arguments in this area and present them in their generic form, without much commenting or discussing at first. I think this is very important for understanding correctly what are the original claims in those arguments. In focusing on the details of some the central arguments in this area I develop my own position and arguments steadily along the way. That is why I don't offer any substantial criticism of the anti-physicalist argument until chapters three, four and five. I begin developing my own position in chapter three, where I present the phenomenal concept strategy and I discuss several issues that are of central importance for the proper understanding of psychophysical identity. Finally in the chapter five, I analyze the vagueness constraint for the phenomenal concepts and the role a study of psychological concepts can have in understanding and explaining away the intuition of distinctness.

In this thesis, I have tried especially hard to avoid long historical expositions of the problems or arguments because the key problems in this area did not change over time in a way that would significantly affect our understanding of them. Historical or chronological considerations can often be very useful for understanding of some problems but not in the case of the explanatory gap problem. I rather focus on contemporary debates and arguments and sketch the theoretical positions directly from there. For example, the mere fact that Frank Jackson has changed his mind about the consequences of the Knowledge Argument (Jackson 1982) adds nothing to the plausibility or persuasiveness of his new position. In fact, different persons might have come up with the arguments, and the arguments would remain as plausible as they are now.

However, the fact that I kept philosophical/conceptual issues apart from the empirical issues until the very end should not prevent me from stating briefly at this point what are my general ideas about the relation between what neuroscience can say about consciousness and the explanatory gap problem.

I argue that any theoretical resources from neuroscience cannot directly solve the explanatory gap problem because of the constraints encapsulated in the notion of vagueness

of phenomenal concepts[5]. Namely, given that phenomenal concepts[6] refer directly, not via descriptions of the causal roles, and thus cannot have a priori connections with the material properties, we should conclude they are vague. This means that we cannot use any sort of empirical methods to study phenomenal consciousness directly. A proper way in which empirical evidence from neuroscience, cognitive science and psychology can help us solve the explanatory gap problem is through studying "psychological concepts" in the framework of quality space. The psychological concepts are concepts that use descriptions of the functional roles as modes of presentation and the details of their workings are not necessarily set in the same way as in the natural kind concepts. I propose to analyze them in quality space models in order to unveil why phenomenal concepts are expected to refer via descriptions of the causal or functional roles. The expectation that phenomenal concepts should refer via descriptions of the causal, or the functional roles, is what I think lies at the heart of the "intuition of distinctness" (Papineau 2002). Given that phenomenal concepts are vague in their very nature and that there is no available scheme of the descriptions of their causal or functional roles I argue that we should not expect to find an answer to the central question of the explanatory gap problem, i.e. why certain phenomenal states are associated with certain neural states or any states at all for that matter, directly in the psychophysical identity, as some philosophers think we can in the case of natural kinds (Chalmers and Jackson 2001). I urge that we need an independent account, external to the phenomenal concept semantics to explain this situation. Such an account is the quality space model. The quality space should be understood here as a multidimensional space consisting of several axes of relative similarity and differences among the structures of ordering in different modalities of conscious experience. The important thing about the quality space is to discern which axes and which modalities should be represented in the quality space. On my proposal it is possible that some axes in the quality space consist of their own quality spaces so we could "zoom in" and "zoom out" into the descriptions of the functional roles and see more

[5] This will be a major topic in the fifth chapter.

[6] Concepts of consciousness from the first person perspective, or concepts of what it is like to undergo certain experience.

clearly what the explanation of certain aspects of consciousness looks like when thought of in terms of psychological concepts.

Dissertation's structure

Now it is time to say briefly what my topics will be and what will be going on in each chapter of this dissertation. Apart from that I also need to say why I have chosen to discuss the arguments and problems presented in the chapters.

In the first chapter I start off with Nagel's (Nagel 1974) account of the epistemic gap (section 1.1). His account is very important for the discussions in philosophy of mind in general and especially for the debates about the explanatory gap problem because he introduces for the first time a notion "what it is like" of experience which later on transforms into a technical term "qualia"[7]. His paper "What is it like to be a bat" (Nagle 1974) plays a central role for two main reasons: the first is that he introduces the epistemic gap, and the second is that he provides a version of a proto-knowledge argument.

The next section (section 1.2.) in the first chapter is on the Knowledge Argument (Jackson 1982). As already mentioned earlier the basic idea behind the Knowledge Argument is that we cannot know or infer phenomenal facts from the physical facts, because the phenomenal is epistemically very peculiar. It is accessible only through the first person perspective. Jackson provides three thought experiments to corroborate this claim, one of which includes the now famous Mary, a brilliant neuroscientist who is confined in a black and white room and knows everything physical (biological and neurological as well) there is to know about colour vision but still learns something new after being released.

In the section 1.3. of the first chapter I present Kripke's modal argument against materialism (Kripke 1980).

Blumenfeld formulates Kripke's argument as follows:

[7] "Qualia" is a plural. The singular version of the term is "quale".

“8. Identity statements whose terms are both rigid designators have to be necessarily true in order to be true.

9. 'My pain' and 'my brain's C-fibre stimulation' are rigid designators.

10. Therefore, 'My pain is my brain's C-fibre stimulation' has to be necessarily true in order to be true.

11. But 'My pain is my brain's C-fibre stimulation' is not necessarily true.

12. Therefore, 'My pain is my brain's C-fibre stimulation' is not true.” (Blumenfeld 1975, p. 151)

An important point in Kripke’s argument is the inference from conceivability to metaphysical possibility, which will play the central role in Levine’s account of the explanatory gap. This is the main reason I have chosen to consider his argument in the first place. The entailment from conceivability to metaphysical possibility is also the common target for the defence of physicalism.

In section 1.4. and its subsections 1.4.1, 1.4.2 and 1.4.3 in the first chapter I discuss the zombie argument (subsection 1.4.1), its structure and notions of supervenience and necessity employed in it (subsection 1.4.2) and finally the argument from supervenience for the ontological gap (subsection 1.4.3) by David Chalmers (Chalmers 1996). Both the zombie and the supervenience arguments are very important because they try to justify the entailment from conceivability to metaphysical possibility in the modal argument on the ground of our modal intuitions and to establish property dualism. In this section I also discuss varieties of responses to the zombie and the supervenience arguments and I briefly portray other positions in philosophy of mind. I do that by considering what consequences we get when denying particular premises in Chalmers’s arguments. In this section I also discuss different understandings of necessity which are essential for the debate about the explanatory gap problem.

In the section 1.5. I present one of the most technically complicated accounts in this area, namely, the two-dimensional (2-D) argument (Chalmers 1996, 2009). Chalmers’s 2-D account is tremendously important for the defence of conceivability arguments because it is

the most elaborated attempt to defend entailment from conceivability to metaphysical possibility. I rehearse the zombie argument and the modal argument in the 2-D framework. In that section I also discuss a distinction between primary and secondary intensions which is very important for understanding Levine's account of the explanatory gap (Levine 2001).

Chapter two is completely devoted to laying out the explanatory gap account (Levine 1983, 1993, 2001). In the section 2.1 I discuss the key notions such as thick and thin concepts and modes of presentation. These issues are essential for Levine's account of the explanatory gap and for the phenomenal concepts, which I discuss in the third chapter. The issue of modes of presentation is central to my overall approach, as I will be arguing in the fourth and fifth chapters. In the section 2.2 I discuss a relation between ascriptive and non-ascriptive modes of presentation and their relations with apriority and necessity. In the section 2.3 I discuss notions of "thick " and "thin" conceivability. These issues are very important for Levine's account to work because they are supposed to stop the inference from conceivability to metaphysical possibility. In the section 2.4. I examine notions of "gappy" and "non-gappy" identities as a direct consequence of the considerations about thick and thin conceivability. The next issue that follows it is the intelligibility of explanations which I discuss in the section 2.5. This issue plays a prominent role in my discussion and will be one of the major topics in chapters four and five. Finally in the last section 2.6. I examine some problematic issues with Levine's account of the explanatory gap and set the stage for the next chapter which is about the phenomenal concept strategy.

The first section (section 3.1) in the third chapter is about Loar's account of the recognitional phenomenal concepts (Loar 2002). I examine some of the main points of his account. I have chosen to start with Loar because he lays out a foundation for the whole strategy by setting phenomenal concepts aside from other sorts of concepts. One of the main tenets of this approach is that phenomenal concepts refer directly so are irreducible in that sense but, on the other hand, guarantee a posteriori necessary status of psychophysical identity. The next section 3.2 is about Tye's account of phenomenal concepts (Tye 2003). I chose Tye's account because it nicely develops the other main features of phenomenal concepts and provides a good criticism of the early Loar's account. In the section 3.3 I present and discuss indexical-quotational and perceptual varieties of phenomenal concepts. The perceptual variety is, in my opinion, the most advanced account of phenomenal

concepts. In the section 3.4 I discuss the *inconsistency objection* to the Knowledge Argument and argue that one cannot hold phenomenal properties causally inefficacious and epiphenomenalism true. If they were causally inefficacious we would not be able to detect them in the first place. The main idea of this section is to argue for the first premise of the causal argument[8]. Section 3.5 and its subsections 3.5.1 and 3.5.2. is about some of the most difficult issues of the phenomenal strategy. I start in subsection 3.5.1 by examining the so-called ”Max Black’s objection” and associated issues of modes of presentation. In the next subsection 3.5.2 I discuss another really important issue with the phenomenal concept strategy, that is, whether phenomenal concepts are physically explicable and if they are whether they can explain the explanatory gap.

In chapter four I develop my key argument. In the section 4.1 I rehearse the arguments for the ontological gap in order to be able to pinpoint my main points. In section 4.2. I discuss intuitions underlying the conceivability arguments and the explanatory gap. In the next section 4.3 I provide another take on the issue of intelligibility of explanations. In sections 4.4 and 4.5 I analyze the brute disagreement about tacit assumptions in the conceivability arguments and the explanatory gap. Finally, in the last section 4.6 I discuss the role of a priori derivation in intelligibility of explanations.

The main topic of the fifth chapter is vagueness of phenomenal concepts and ways in which voluptuous empirical research can be used in solving the explanatory gap problem. In the first section 5.1 I discuss vagueness constraint of phenomenal concepts. In the section 5.2. I analyze intuition of distinctness in the light of vagueness constraint. In sections 5.3 and 5.4 I examine Clark’s (Clark 2000) and Rosenthal’s (Rosenthal 2010) accounts of a quality space that is of central importance for studying intuition of distinctness. Finally, in the section 5.5 I present my account of the quality space for pain.

In the last chapter six I give concluding remarks and sum up my whole approach.

[8] That is: mental causes have physical effects.

CHAPTER 1

1.1 Arguments against physicalism

In this chapter I shall examine several classical arguments that try to establish either the falsity of physicalism or its inherent epistemological inadequacy. The objective of these considerations is to set up a stage for the analysis of the explanatory gap problem. It is of crucial importance to clear up the basic terminology, as well as the underlying assumptions in the arguments in order to be able to understand properly the explanatory gap. To do so, I need to rehearse arguments that are the basis for the ontological gap positions[9], as well as the account of the epistemic gap position (Nagel 1974). I won't be giving any comprehensive criticism of these arguments at this stage because I want to present the original arguments in their pure form. This tactics is very important for my overall approach because I am developing my own position steadily throughout the chapters and directly from the analysis of the arguments. Chapters three, four and five are all about my own arguments and position regarding these problems. Chapter two is exclusively dedicated to analyzing Levine's account of the explanatory gap (Levine 1983, 1993, 2001).

[9] Kripke's modal argument (Kripke 1980), Chalmers's supervenience argument (Chalmers 1996) and two-dimensional argument (Chalmers 2009) and the Knowledge Argument (Jackson 1983).

1.2 Nagel's Account

We are in the same position towards psychophysical identity as ancient Greek philosophers would have been towards the claim that matter is energy, although it is true, they could not understand it, for the conceptual reasons

Let's begin with the Nagel's account. In its core the explanatory gap problem is based on the claim that the "what is it like" aspect of experience (Nagel 1974) is left out from an empirical explanation of phenomenal consciousness (Levine 1983, 2001). According to this view there is an unbridgeable gap in our understanding of phenomenal consciousness, because it is assumed that any empirical theory of phenomenal consciousness leaves out qualia[10], which in effect makes the psychophysical identity unintelligible even if it were true.

In order to clarify all of these claims and terms let us take a closer look at arguments corroborating them and then try to pinpoint their common underlying assumptions. Strategically, the best starting point would be an explanation of the "what is it like" aspect of experience. This term has been introduced by Thomas Nagel in his now classical paper "What is it like to be a bat" (Nagel 1974). Nagel famously claimed that there is an unjustified use of analogies and paradigms from empirical sciences, especially the computer or artificial intelligence paradigm, in the reductive strategies of explaining consciousness. What makes them unjustified is that the attempts of psychophysical reduction through identity of the type H_2O=Water, or comparison of the mind and brain with the hardware and the software, do not take into account that what is to be reduced is unique in many ways. On his account,

[10] "Qualia" or "quale" as a singular, are technical terms denoting the "what it is like" aspect of experience.

although physical structure or functional organization may vary from organism to organism, the very "fact that an organism has conscious experiences at all, means that there is something it is like to be that organism" (Nagel 1974, p. 436). He calls this feature "the subjective character of experience". We may put it more formally and clearly in the following way:

Subjective character of experience **(SCE)**: An organism has conscious mental states iff there is something that it is like to be that organism, for that organism (*Ibid, p. 346*). He claims that the absence of SCE is logically compatible with all the reductive analyses of the mental in his time, because it cannot be captured by any of them. When considered in terms of contemporary taxonomy, Nagel's move can refute only type A materialism (Chalmers 2003), as its proponents claim that there is only one kind of property in the world and therefore only one kind of concept to pick them out. However, as we shall see in the chapter on the phenomenal concept strategies, this claim should not worry a type B materialist (Ibid.) because they may also maintain that there is only one kind of property in the world that is picked out by two distinct modes of presentation[11].

This point alone is the most important for the knowledge argument because on a different reading it says that undergoing certain experiences is a necessary condition for directly conceptualizing them and, as we shall see, it's also very important for the conceptual dualist account of phenomenal concept. This argument is based on the claim that SCE is not analyzable in terms of functional or intentional states because we could ascribe mental states to the robots or automata that behaved like humans although experienced nothing. In other words, any form of the explanatory system in terms of functional states does not allow for a differentiation between creatures with SCE and creatures without SCE. On the other hand, Nagel claims that if there could be robots complex enough to behave like a person they might have SCE, but their SCE *could not be* discovered merely by analyzing the concept of experience (Nagel 1974, p. 436, footnote 2). He goes even further in claiming that SCE is not analyzable even in terms of the causal role of experiences in connection to typical human behavior for the same reasons. Therefore, he concludes, without an account of what

[11] Cf. (Papineau 2002).

the subjectivity of experience is, we would not know what is required for a physicalist theory of consciousness. In order to elaborate this idea let us look back at the difference between the body-mind problem and the examples of successful reductive explanations in natural sciences mentioned at the beginning of this section. The problem with the SCE is that SCE features cannot be excluded from the reduction in the same way as these features can be excluded from the reductive explanation of the ordinary, objective/empirical phenomena. For example, certainly it is like something to see the rainbow or to feel gravity pull, i.e. one's own weight, but this feature of the experience of rainbow, say, does not figure anyhow in the physical explanation of rainbow, because it can be eliminated from the explanation just by showing what effects the actual rainbow has on our minds. This fact that the SCE is bonded to the point of view amounts to the difference between the conception of subjective and objective which is fundamental to the problem. In another words, *to observe SCE is to have SCE*. If there is no intermediary epistemic instance between the observer and the phenomenon then it is not clear how we can overcome the conceptual barrier, because the SCE is not objectively available, i.e. it is not available from the third person perspective.

To illustrate this idea more vividly Nagel introduces an analogy with bats. They seem to be suitable candidates for the analogy, because they are phylogenetically close enough to us, humans, but with a fundamentally different sensory system. They perceive the surrounding world by echo-location, so it is difficult to compare their phenomenology with our phenomenology. Although we could ascribe some kind of mental states to them, based on what we know about the physics of sound waves, namely, knowledge of how sonar works as well as some neuroscience concerning both human and bat brains, it seems clear that what it is like subjectively to be a bat is nothing we can experience or imagine. Even *if* we were able to imagine, in staggering detail, what it would be like to be a bat, it would only help us understand what it would be like *for us,* as humans, to be a bat. Nagel makes this claim more firmly by suggesting that even if it were possible to suddenly or gradually alter our physical and neurological structure it is hard to extrapolate what it would be like to be a bat. This analogy brings us to the more formal epistemic claim about the relation between facts and conceptual schemes or systems of representations, as Nagel (Nagel 1974, p.441) puts it. This claim, that there are facts of which humans will never possess respective concepts or which will remain incomprehensible, simply because our physical structure does not allow us to operate with concepts of the requisite type, is of central importance to the

account. This is not the state of affairs merely because our ability to comprehend it is limited by our finite nature. For example, transfinite numbers would have still existed and they would have been comprehendible even if there were no one to discover them or to comprehend them. This is so because the *ability* to comprehend certain facts and to utilize them conceptually in a theory is ontologically tied to the respective point of view**.** Furthermore, to have this unique and ineffable point of view is to have a respective physical structure. That is the reason why humans will never know what it is like to be a bat, according to Nagel. However, it is also important to note that our inability to comprehend facts that are beyond our conceptual grasp does *not* prevent us from being able to state or recognize their existence. Let us call this thesis the `*Experiential Thesis'* (ET). Nagel explains it more clearly in the footnote (Nagel 1974, p.442, footnote 8). He says that ET, as formulated above, is not merely the epistemological issue but rather a more fundamental issue according to which even to form a conception of the SCE of a given creature, one must adopt a point of view of that creature. And if physicalism is true, taking up a point of view requires having a respective physical or functional structure. This means that we can know something about the SCE of a relevantly similar creature, just by drawing analogy with our own experience, for example. This does not seem to be a problem. However, a problem does emerge when one tries to come up with a conception of the SCE which is not her/his own. As noted from the ET, to have a conception of the relevant SCE one must assume the point of view of that SCE. If this is the case, then SCE cannot be studied objectively because the facts, and therefore conceptions, about SCE are only subjectively accessible.

However, Nagel claims that this is not an argument against reduction in general, because for example a Martian scientist with no knowledge or understanding of visual perception could be able to understand lightning or the rainbow, because understanding these phenomena does not require the ET. This is so because the concepts that we use to pick out things like lightning or the rainbow are tied to a particular point of view, but the objective nature of those things is external to any point of view (Nagel 1974, p. 443). For example, lightning has an objective character that is not exhausted by its visual appearance, whereas the SCE *is* the appearance itself. The truly successful reduction can only work if the species-specific viewpoint is omitted from what is to be reduced (Nagel 1974, p. 445). This is one of the central theses of Nagel's argument. According to this thesis the what-it-is-like aspect of experience is intrinsic to the experience. Nagel concludes that physicalism is most

probably true, because its central hypothesis is rather clear: mental states are part of the physical world and therefore they must be explicable in empirical terms. The only thing that we lack at the current level of understanding is how this hypothesis could be intelligible. Nagel uses another analogy to illustrate this point: we are in the same position towards the thesis mental states are brain states as pre-Socratic philosophers would have been towards the thesis that matter is energy, the latter thesis is true but the pre-Socratics lacked the relevant conceptual framework to fully understand the thesis.

Now, the central problem of this argument is that the concepts of an objective/empirical theory of consciousness cannot refer to the phenomenal properties because a conception of the phenomenal properties is only possible through direct acquaintance or undergoing of them, whereas objective conceptions operate independently from an epistemic perspective. A further step towards articulating the explanatory gap would be to claim that given the fundamentally different ways phenomenal and physical concepts work, an empirical theory of consciousness leaves out the what-is-it-like aspect of consciousness, and therefore there is the explanatory gap.

This is one of the crucial points that underlies the knowledge argument and the conceivability argument. I elaborate on this point below.

The knowledge argument relies on this assumption thereby claiming that knowledge of all the physical facts does not entail knowledge of the phenomenal facts, for if it did, then Mary would not have learned anything upon her release. However, this point can be read in a slightly different way. It might mean that if a conception of the phenomenal is experientially dependent then of course it is not contained or inferable from the body of physical information. If this is so, then the only problem we have left is to show how *phenomenal concepts* might pick out *physical properties*, not the other way around, because if phenomenal concepts are different from physical concepts, then it should not be expected that they pick out their referents by the causal role they play but rather in a direct way[12]. In this manner, we can concede that there are two kinds of concepts, which are picking out the same properties, one in terms of the causal roles the properties play and the other one in

[12] For the details of this account see Papineau (2002).

terms of their experiential quality. Let's follow Papineau (Papineau 2002) and call this position "conceptual dualism". This point will be elaborated in much greater detail when we come to the exposition of Papineau's account of phenomenal concept in the following chapters.

The conceivability argument also hinges upon Nagel's point in a very important way. Namely, because of the way phenomenal concepts pick out their referents, a misdescription as a way of explaining away the seeming contingency of the psychophysical identity is precluded. What appears as pain, is pain. We will come to this point very shortly and see how the same conceptual dualism strategy applies to the conceivability argument.

1.3 The knowledge argument

Phenomenal facts are neither contained nor inferable from the body of physical information

Although the basic idea behind the explanatory gap problem seems very simple, i.e. we cannot understand how phenomenal consciousness could be explained in terms of neuroscience, it needs to be refined in order to see what makes it so compelling. After having introduced one of the basic claims upon which the explanatory gap problem relies, that is the idea of leaving out the "what is it like" of experiences from the physicalist explanation of consciousness, we need to move on and examine the arguments that start with this premise and lead to ontological claims about phenomenal properties. Therefore, the next step would be to present and analyze the famous Knowledge Argument (Jackson 1982, 1986).

The starting premise of Jackson's argument is the claim that no amount of purely physical information could say anything about qualia. His reading of the notion "physical information" is supposed to be central to the argument and, as we shall see, to the understanding of the explanatory gap problem.

He develops three different arguments in order to show how his own account fit in with the "qualia freak" intuitions as he calls them, i.e. intuitions that qualia are left out by any kind of physical or neuroscientific theory of consciousness.

The Knowledge Argument, as it is commonly known, relies on the claim that not only are qualia not included in the physical information about the universe, but the information about qualia is not even inferable from the physical information.

He develops two thought experiments to corroborate this claim. The first one involves a person, Fred, who is able to discriminate an additional shade of red color that no one else except him can discriminate. Jackson calls them red1 and red2. To Fred, say red2 is so distinct from red1 as yellow and blue are distinct to all the other people. In a sense, it is not just a shade of red color to Fred but a completely different color. Furthermore, he also has different sensory and phenomenal experiences of red2. He has tried many times to teach his friends about the different shade of red but he always failed because everyone else lacks both sensory and phenomenal discriminatory ability. In a way, Fred is in a same position as a sighted person within a blind community; he can never convince the others that he actually has another sense. Although he avoids pits, ditches and all kinds of other obstacles, for the blind community he is just a skilled blind man. In a sense, Fred's experiences of red2 are ineffable.

However, the question immediately arises: *what kind of experience Fred has when he sees red2.* What is it like for Fred to see red2?

Imagine after Fred's death, his visual system, including eyes, optic nerve, and the whole brain, were preserved, and some kind of keen neuroscientific research group wanted to answer these questions by performing all kinds of empirical experiments on it. First, they discovered that there are some additional cones in Fred's retina that respond to the light of the wavelength that was assumed to correspond to that of red2. Or that he has normal cones, but they respond to certain wavelengths that make no difference to the cones of a normal perceiver. Now imagine this research group was so advanced that they were able to replicate Fred's physiology in everybody else, or to transplant his optical system to someone else. They performed thorough tests and experiments on people with replicated physiology

and transplanted optical system but no one in both groups could have discriminated red1 from red2.

Here comes the crucial point of the thought experiment: how is it possible that they knew all the physical information about Fred, and if physicalism is true, they should have known that *ex hypothesi,* and yet, they could not know what is it like for Fred to discriminate between red1 and red2?

The answer imposes itself for Jackson. Physicalism is incomplete for it leaves something out (Jackson 1982, 1986).

Before trying to formalize this argument let us have a look at another take of this argument. This is of course the brilliant neuroscientist Mary thought experiment.

The scenario includes Mary, a neuroscientist who learns about the neurophysiology of color vision while confined in a black and white room, through black and white television and books. Of course, there is nothing wrong with her visual system, and she learns about wavelengths of light and about retina and optic nerve, about utterances of color notions, she also learns that sky is blue, grass is green and roses are red. The point is that she learns everything there is to know about physiology of color vision without being able to actually experience color as everybody else. Now Jackson asks a question: what will happen when she is released from the black and white room (Jackson 1982, p 130)? Will she learn something new after finally seeing the colors? Jackson claims that it seems obvious that she will learn something new from her own newly acquired experiences. If she learns something new, Jackson concludes, then physicalism is false (ibid.). This conclusion is based on the premise that if physicalism is true then it is not possible to have all the physical information about color vision without having all the information about color vision. If she learns something new then having all the physical information does not include all there is to know about color vision. But *ex hypothesi* she should have known everything there is to know about color vision. Therefore, what she learns from her own first person perspective, her qualia, is left out from the physical information she acquired beforehand. Of course, the same argument applies to other bodily sensations that are said to have qualia, like olfactory, gustatory, auditory, etc.

The argument is formalized as follows:

"13. Mary (before her release) knows everything physical there is to know about other people;

14. Mary (before her release) doesn't know everything there is to know about other people (because she *learns* something about them on her release);

Therefore,

15. There are truths about other people (and herself) which escape the physicalist story" (Jackson 1986, p. 293).

One of the responses to this argument was that Mary did not learn a new fact, but that she learned the old facts in a new mode of presentation, namely from her own first person perspective (Churchland 1985, Jackson 1986). Since *ex hypothesi* she should have known everything about colour vision if physicalism was true, it seems rather obvious to Jackson that whatever knowledge she gains, it is not about anything physical. This line of reasoning seems impenetrable. However, nothing in this argument compels us to conclude that whatever new she learns is based on a metaphysically distinct property. As we shall see in the third chapter, the Knowledge Argument actually helps to establish conceptual dualism, a doctrine which acknowledges the epistemic gap but denies that there are two metaphysically distinct kinds of properties.

In order to fortify his account further, Jackson analyzes his own argument through the so-called modal argument (Jackson 1982, p 130). It is claimed by the modal argument that no amount of physical information would logically entail knowledge of other minds. Given this premise then there is no contradiction in conceiving a possible world in which creatures exactly like us, humans, in all functional and physical respects, would not have mental states, in other words, they would lack qualia. If we and our functional and physical duplicates were alike in all respects but they lacked something that we have in addition to functional states and physical history, then whatever they lacked is not functional or physical. Again, this argument holds because *ex hypothesi* if physical information explains everything there is to be explained about sensory states and mental states, then the

possibility of such creatures would have been a contradiction. Hence, physicalism is false (Ibid.).

In order to back up the modal argument from the objection that physicalism is envisaged to be a contingent theory, so that modal intuitions should be restricted to the actual world and the similar possible worlds, Jackson points out that even in such a restricted modal topography the knowledge argument holds, because even knowledge of other minds is not entailed from the physical information.

At this point it is very important to note the difference between Nagel's "what is it like to be" account (Nagel 1974) and the Knowledge Argument. Although both arguments make very similar claims, i.e. that qualia are left out from the physical story, Nagel argues that if we come up with a theory of the phenomenal we might be able to know what is it like for others to undergo or to have mental states, so although it would not help us to overcome the point of view barrier we would know more about qualia and still maintain physicalism as true, whereas Jackson claims that physical information does not provide even logical entailment of qualia. Clearly, Jackson's argument makes a more severe impact on physicalism because it follows from it that even if we would have developed some theory of the phenomenal we would never know more about mental life in terms of physical information. Of course, physicalists cannot concede on this because they actually want to explain qualia in terms of physical information.

Now, in order to reconcile materialist intuitions with the severe consequences of the knowledge argument, Jackson embraces epiphenomenalism, an ontological position according to which phenomenal properties are totally functionally and causally inefficacious towards the physical properties, but they may be efficacious towards other phenomenal properties. Recently Jackson has assumed more consequent ontological position (Chalmers and Jackson 2001), which is a consequence of the knowledge argument, a form of dualism, i.e. the property dualism, but we shall come to this point later on when discussing the two-dimensional modal framework. Let us now work out reasons for assuming epiphenomenalism on the ground of the knowledge argument.

In arguing for the epiphenomenalism Jackson makes it very clear why it is the only reasonable position if we want to maintain physicalism as a true theory and still being

confronted with the devastating consequences of the knowledge argument. Namely, if qualia are totally functionally and causally inefficacious towards physical properties and physicalism is true, then qualia must be some by-product of certain brain processes that are efficacious towards physical properties. To illustrate this point he analyzes a physical trait of having a warm and heavy coat for the survival in the Arctic within a theory of evolution. According to the theory of evolution, say, polar bears have developed a thick coat to be able to survive in the very cold polar environment. Now, a thick and warm coat is also heavy, so one might analyze the evolution of the polar bear's coat in such terms as if the heavy coat was conductive for survival (Jackson 1982, p. 134). This line of thinking seems obviously wrong, because it is the thick and warm coat that is conductive for survival, and its heaviness is a by-product that actually slows the animal down, and thus is not conductive for survival. However, benefits of having a thick and warm coat in evolutionary terms outweigh disadvantages of having a heavy coat. Having this analogy in mind, Jackson claims that we are in a same position towards qualia as we are towards heaviness of the polar bear's coat. Qulia, although functionally and causally inefficacious, are epiphenomena of some efficacious brain processes. Jackson's point here is if we try to analyze phenomena in the terms of causality, it might turn out that we connected wrong causes and effects in the first place. He concludes the that qualia do not cause anything physical, but are caused by the physical (Jackson 1982, p. 134)

1.4 Kripke's modal argument

Identities if true at all, must be necessarily true. Psychophysical identity does not seem to be necessary true; therefore, it is necessary false.

In order to see more clearly how the conceivability arguments pose problems for physicalism and get some handle on modal issues surrounding the explanatory gap problem, we need to rehears Kripke's modal argument in some detail.

Kripke (Kripke 1980) says that identity statements if true at all, must be necessary true. This means that terms flanking the identity sign must pick out the same thing in every situation, imaginary or actual. That is to say, they must pick out the same thing at all possible worlds[13]. According to Kripke, terms that pick out the same thing in all possible worlds are called "rigid designators". They are called "rigid designators" because they refer rigidly-i.e. they pick out the same thing at all possible worlds. The original class of rigid designators comes from Kripke's theory of proper names. In criticizing descriptivist theory of meaning, which I won't be discussing here for the sake of avoiding unnecessary detours, Kripke argues that proper names don't refer via descriptions, but directly. The reference in this case is fixed by an act of "baptism". The natural kind terms, in his view, are also rigid designators as well as terms such as "pain". Although he treats natural kind terms as names, he maintains that basically in that case names are equivalent to descriptions. The reference of natural kind terms is fixed by a contingent property that we causally associate with term, which guaranties its *a posteriori* status. However, what guaranties the necessary status of identities is not a direct link between two names, but rather a link between an object and itself.

In Kripke's view, some identity statements seem to be contingent, we can conceive of their falling apart. For example, it seems that in the case of statements like "Water is H_2O" or "Heat is molecular motion" something else than H_2O or molecular motion could have been water and heat respectively. This pattern of reasoning should be applicable to cases such as "pain is the C-fibre firing". After all, both kinds of cases are supposed to be *a posteriori* necessary. But if such dissociations are conceivable then these terms pick out different things at different worlds. If this is so, then the identity statement is not necessary according to Kripke. However, he claims that this is merely apparent contingency, because identity statements like "Water is H_2O" or "Heat is molecular motion" use rigid designators, thus must be necessary. In the case of "water" and "heat" the apparent contingency is explained away as a misdescription. What we are conceiving of when thinking about statements like "heat is not molecular motion" is not that something else than the motion of

[13] The notion "possible world" denotes ways in which the world might have been. This includes the actual world as one of the possible worlds, but in that case it is centered. I won't be engaging in discussion on different views about possible worlds in this project, simply because it is out of scope of this work and it would not change much in my discussions.

molecules is heat, but rather that something else than the motion of molecules produce a sensation of heat. Indeed, there could be creatures so different from us whose sensory system reacts to something else than motion of molecules as heat. But if "heat is molecular motion" is necessary, then it is not conceivable that something else than molecular motion could be heat, because both "heat" and "molecular motion" are rigid designators and as such must always pick out the same thing. In this case we are conflating a term that is really contingent, i.e. that something else than the motion of molecules could have produced a sensation of heat, with the necessary statement that heat is molecular motion.

On the other hand, the properties these descriptions invoke, are only contingently associated with the terms. For example, the reference of "heat" is fixed by a description that we treat as a name: "molecular motion", but it has a property of causing heat sensation only contingently, that is to say that something other than molecular motion could have been causing the sensation of heat. Given that properties are only contingently associated with the term, it allows for a misdescription of a counterfactual situation in which a property does not satisfy a description. For example, a world in which heat is not molecular motion does not seem conceivable, whereas in fact this statement only claims that something other than molecular motion could have caused a sensation of heat.

On this account natural kinds terms[14] refer via a contingent property because it might have turned out that water were not H_2O or that it had some different macro-physical properties than odourless, transparent, etc. However, since it might have turned out otherwise, Kripke argues, although *a posteriori*, it is a necessary identity, because then the whole world would have turned out otherwise. The fact that terms refer via a contingent property is essential for Kripke's argument against materialism, because it actually makes these *a posteriori* identities appear contingent whereas in fact they are necessary and the sense of contingency is explained away as a misdescription.

The argument actually starts from the claim to which materialism is committed: if true at all, psychophysical identity must be necessary true, in other words, a situation in which it comes apart should not be possible. Then the first claim comes into play and states

[14] Terms like: water, heat, molecular motion, H_2O, mean kinetic energy, etc.

that the psychophysical identity's coming apart is conceivable. Note that in the case of natural kind identities first premise is the same, it is conceivable that something other than molecular motion causes sensation of heat, because concept of heat refers via contingent property, so it might have turned out that really something other than mean kinetic energy were causing heat sensations in humans, but it is not possible that heat is not molecular motion. However, as we discern that truly necessary and contingent statements are conflated here, we classify it as a misdescription and conclude that actually such a situation is not even conceivable. Such a pattern of analysis is not available for the psychophysical identity.

Kripke argues that this pattern of analysis is not applicable to "pain=C-fibre firing" case. If we assume that "pain=C-fibre firing" is an a posteriori necessary statement of exactly the same type as "heat" and "water" cases, then "pain" and "C-fibre firing" have to rigidly designate as well. However, argues Kripke, there is no contingent property that fixes the reference of "pain" like in the case of "heat" or "water", so what appears contingent in the case of "pain" is in fact contingent. In the case of "pain" we cannot account for a misdescription in explaining away it seeming contingency. So if "pain=C-fibre firing" is conceivably false, it is false. Therefore concludes Kripke, "pain" and "C-fibre firing" cannot be identical.

As we recall Blumenfeld's formalisation of the argument:

"8. Identity statements whose terms are both rigid designators have to be necessarily true in order to be true.

'8. My pain' and 'my brain's C-fibre stimulation' are rigid designators.

10. Therefore, 'My pain is my brain's C-fibre stimulation' has to be necessarily true in order to be true.

11. But 'My pain is my brain's C-fibre stimulation' is not necessarily true.

12. Therefore, 'My pain is my brain's C-fibre stimulation' is not true." (Blumenfeld 1975, p. 151)

It seems reasonable to assume that for Kripke conceivability entails possibility. To put it in more technical terms, for Kripke conceivability entails metaphysical possibility. So if an identity statement containing rigid designators is conceivably false, it is necessary false. If "My pain≠C-fibre stimulation" is metaphysically possible then psychophysical identity is necessary false.

The common strategy in attacking Kripke's modal argument is denying the inference from conceivability to metaphysical possibility. For instance, Christopher Hill (Hill 1995) urges that when we are conceiving of psychophysical identity's falling apart we employ two very different and distinct imaginative faculties. Hill alludes to Nagel's (Nagel 1974) often neglected point about perceptual and sympathetic imagination. Nagel claims that when we are imagining something perceptually we put ourselves in a state resembling the one in which we would be when perceiving it. On the other hand, when imagining something sympathetically we put ourselves in a conscious state resembling the thing itself (Nagel 1974, pp. 445-6, footnote 11). Nagel explicitly writes: "Where the imagination of physical features is perceptual and the imagination of mental features is sympathetic, it appears to us that we can imagine any experience occurring without its associated brain state, and vice versa. The relation between them will appear contingent even if it is necessary, because of the independence of the disparate types of imagination." (Nagel ibid.) Now, Hill (Hill 1995) and Hill and McLaughlin (Hill and McLaughlin 1999) claim that even if perceptual and sympathetic imagination independently from each other were reliable guides to conceivability their joint exercise can be extremely unreliable. Hill calls this sort of conflating "Cartesian intuitions". He maintains that Cartesian intuitions are not reliable guide for the entailment from conceivability to metaphysical possibility.

On the other hand, Joseph Levine (Levine 1983, 1993, 2001) argues that the notion of conceivability in Kripke's argument does not justify inference to metaphysical possibility. He says that the conceivability arguments only manage to establish epistemic possibility. Levine thinks he saves physicalism in this way but at the price of an explanatory gap. I will be discussing Levine's account of the explanatory gap in the next chapter, so I leave the comments for then.

In response to this kind of criticism David Chalmers (Chalmers 1996, 2009; Chalmers and Jackson 2000) offers much more elaborated accounts of the modal argument (zombie argument) and of the inference from conceivability to possibility (two-dimensional argument). In the rest of the chapter I will be discussing Chalmers' accounts of the argument and the inference from conceivability to possibility.

1.5 The argument from supervenience for the ontological gap

If the phenomenal is not logically supervenient on the physical, and the physical domain is causally closed, then phenomenal properties are distinct properties from the physical properties.

We can think of the argument from supervenience as a two-step argument consisting of two separate arguments. The first argument establishes that phenomenal properties are not supervenient on the physical properties, and the second argument uses this conclusion as one of the premises in conjunction with the thesis about the causal closure of the physical to establish that phenomenal properties are metaphysically distinct properties of the world. The first one is better known as the zombie argument and the second one as the argument from supervenience.

Let's do this step by step. After laying out his supervenience argument and concluding that property dualism must be true, Chalmers argues that natural supervenience better describes the relation between the phenomenal and the physical, and leaves room for the law-like descriptions that are supposed to be discovered by researching the so called *neural correlates of consciousness*. It is not necessary to go into the details of these further claims because rehearsing the core arguments will suffice for setting up the stage for executing my key argument. Let's go now into details of both zombie and argument from supervenience.

The zombie argument can be formalized as follows (where "P" represents a conjunction of all physical facts and "Q" represents any phenomenal fact):

13. (P&¬Q) is conceivable;"

14. If (P&¬Q) is conceivable, then (P&¬Q) is possible;

15. If (P&¬Q) is possible, materialism is false;

---

16. *Therefore*, materialism is false.

17. The supervenience argument can be formalized as follows:

18. The fact that conscious experience exists is a positive fact;

19. Conscious experience is not logically supervenient on the physical;

20. If there are positive facts that are not logically supervenient on the physical facts, then physicalism is false;

---

21. Therefore, physicalism is false and phenomenal properties are further properties of the world.

Let us clarify each one of the steps of these arguments. I shall begin with the zombie argument.

1.5.1 The zombie argument

The first premise claims that it is conceivable that there is an exact physical duplicate of a human being that lacks phenomenal consciousness. The second premise goes further and states that if such a situation is conceivable, it is possible. And finally the third premise establishes that if an exact physical duplicate of a conscious being without phenomenal consciousness whatsoever were possible, then materialism fails. The conclusion in the fourth

step follows by modus ponens. Each one of the three premises in this argument is possible to challenge, and depending on which premise one argues against, one can distinguish among several standard positions in philosophy of mind. It is fairly easy to see what is the weakest point in the argument, which of course has been widely debated. It is the second premise, the entailment from conceivability to possibility. We will come to this issue shortly.

Those who deny the first premise usually embrace eliminativism, or analytic functionalism. Basically, to deny the first premise means to be committed to the claim that phenomenal consciousness does not exist and that the whole zombie paradigm does not make sense, or to claim that consciousness exists but that zombies are not conceivable, thus not possible in any sense. On the face of it, it seems that consciousness exists, and there are only few philosophers who maintain the view that consciousness does not exist (most prominently Dennett 1992, Churchland 1992 and Churchland 1989).

The main line of argumentation assumed by most of the proponents of the eliminativist strategy is to criticize folk psychological characterization of consciousness as misleading. Furthermore, they usually claim not only that folk psychology misdescribes our experiences, but further that the development of neuroscience will remove any sense of puzzlement about how consciousness can be physical.

Analytic functionalism claims that if it were possible to give topic neutral descriptions of the causal roles concepts normally play in a theory then we can a priori translate mentalistic or phenomenal concepts into physical concepts. On their view then, if Mary knew all the physical facts she would have known what it is like to see red colour. The most prominent proponent of this view is David Lewis (Lewis 1972).

Denying second and third premises has become the mainstream strategy. Chalmers calls it type B materialism. What all approaches in this group have in common according to Chalmers is that they all admit the existence of the explanatory gap in one way or another. We will not go into the details of all of them, it will suffice only to present those that are relevant to my key argument. There are several varieties of this strategy, some of which include analyses of the inference from conceivability to possibility purely in epistemic terms (Hill 1997; Ballog 1999; etc.), involving analyses of the notions of conceivability and imaginability as two distinct kinds of cognitive faculties that are employed in conceiving of

zombies, one leading to epistemic and the other leading to metaphysical possibility respectively, but they cannot be used jointly as it seems from Chalmers' outline.

Chalmers, however, argues that those who want to argue against the second premise must be committed to what he calls "the strong metaphysical necessity", which of course he discards as implausible. Chalmers introduces the notion of *the strong metaphysical necessity* (Chalmers 1996, pp. 136-138) as opposed to the *weak metaphysical necessity*, which is a standard Kripkean view on *a posteriori* necessity. The strong metaphysical necessity in Chalmers's terms rests on the assumption that there could be propositions that are conceivable but there is no metaphysically possible world in which they were true. According to this view a proposition is conceivable in accordance with the strongest logical strictures but metaphysically impossible nevertheless. What distinguishes this notion of metaphysical necessity from the standard Kripkean notion of a posteriori necessity is that whereas on the Kripkean view a proposition is considered as conceivable according to its secondary intension[15] but then discarded as misdescription, and thus ultimately impossible; in the case of strong metaphysical necessity, a proposition is conceivable, correctly and coherently described, but still metaphysically impossible. This view seems implausible on the face of it. Chalmers's refutation is based on the claim that such a strong metaphysical constraint would limit a space of possible worlds in such a way that the set of metaphysically possible worlds would be smaller than the set of logically possible worlds, so it would appear that if there are worlds that are logically possible, but metaphysically impossible, then we would never know about them. Chalmers's refutation of this position seems impeccable; however, no one actually argues for the strong metaphysical necessity. Chalmers probably presumes that to deny that zombies are metaphysically possible is to deny that primary and secondary intensions of phenomenal concepts do not coincide and the only way to back up such an assumption would be to embrace the strong metaphysical necessity. As we shall see, an attack to the second premise of the zombie argument does not have to rely on such a bizarre metaphysical assumption.

---

[15] I shall explain the distinction between primary and secondary intensions in a short while, in the section on two-dimensional argument.

One of the most prominent manoeuvres within this approach is of course Joseph Levine's (Levine 1983, 1993, 1998, 2001, 2006) famous explanatory gap account. We will discuss it in much greater detail in the second chapter. Suffice it to say, for the time being, that Levine denies entailment from conceivability to metaphysical possibility, and argues for the entailment from conceivability to epistemic possibility thus allegedly saving the physicalism but at the price of the epistemic gap. We will discuss this position very thoroughly in chapter two. Finally, challenging the second and third premises together presupposes something like phenomenal concept strategy. Let's just mention the main idea behind this approach at this point, as it will be discussed at length in chapter three. Proponents of the phenomenal concept strategy claim that phenomenal concepts are indeed a special kind of concepts whose semantics is based on direct acquaintance with the phenomenal properties or on the direct reference (that does not rely on the descriptions of the causal roles phenomenal properties normally play). The manoeuvre is straightforward with all versions of the conceivability arguments: Mary does learn something new upon her release, but the fact that she learns something new does not mean that she learns about some new, further property of the world, but merely that she acquires a new mode of presentation of an old property. As for the zombies, they claim that zombies are conceivable, perhaps even possible, but their possibility does not put into jeopardy materialism because the possibility of zombies only shows that they haven't acquired a requisite kind of concepts. This is a general outline of the phenomenal concept strategy and it is very crude. It will be discussed in greater detail in its subtleties in chapter three. Phenomenal concept strategy is very important for my key argument as I am arguing from that perspective all along.

Of course, the conclusion in point 35. can't be denied if one accepts all three premises as it follows by modus ponens, one of the fundamental principles of logical reasoning. Basically, if one accepts premise 33., one cannot avoid accepting the conclusion in the step 4. I will be arguing against the premise 33. and I will be defending the perceptual variety of phenomenal concept strategy developed by Papineau (Papineau 2007).

1.5.2 Structure of the zombie argument and notions of supervenience and necessity

Basically, what the zombie argument shows is that the supervenience thesis, according to which phenomenal properties are dependent on physical properties, does not hold. Of course, the argument starts with the thesis to which a materialist must be committed. The materialist's commitment can be expressed in terms of global, local and logical supervenience, according to Chalmers. Supervenience is a dependency relation according to which higher-level properties depend on the lower level properties in such a way that the variations of the lower level properties must be followed by the law-like variation of the higher-level properties. The supervenience thesis in its general formulation reads: there is no difference without physical difference. Or as it is generally put in the debates in philosophy of mind: there is no phenomenal difference without physical difference. That is to say, if materialism, or shall we say the psychophysical supervenience[16], were true, then a situation in which physical properties remained the same whereas phenomenal properties completely were lacking should not be possible.

Now, the superveneince relation can take several forms, depending on the modal or epistemic perspective. If we think about it in terms of possible worlds that determine the meaning of a term, then we have a global supervenience, which formally can be expressed as follows: for any two possible worlds *w1* and *w2* and kinds of properties *A* and *B*: *A* kind of properties supervenes globally on *B* kind of properties *iff* possible worlds *w1* and *w2* do not differ in their *B* properties, then they do not differ in their *A* properties. That is to say *B* properties supervene on *A* properties *iff* the *A* facts determine the *B* facts about the entire world. The "world" here is to be understood as an entire universe, which of course could have turned out otherwise, that is the entire universe corresponds here to a notion of a "possible world".

On the other hand, we have a local supervenience: *A* properties of an individual determine *B* properties of that individual. That is to say: if any two individuals *x* and *y*

[16] Supervenience can take many forms, in aesthetics, economics, etc.

instantiate the *A* properties, they instantiate the *B* properties. In another words, local supervenience on the physical fails when a supervenient property is context or an environment dependant.

It seems rather clear that local supervenience implies global supervenience, but not the other way around. Chalmers's illustration for this claim involves the biological properties that supervene globally on the physical. It seems plausible to assume that two worlds that are indiscernible in their physical properties will also be indiscernible in their biological properties. On the other hand local supervenience might fail in this case because two organisms might differ in some biological properties, for example one is *fitter* than the other, due to some context dependencies, like immediate environment or personal history of that organism. So, if two organisms are indiscernible locally they are certainly indiscernible globally, but not necessarily the other way around.

Chalmers's zombie argument is based on the global supervenience, but it can be applied to local superveneince as easily. This is because of the fact that if consciousness supervenes at all, it supervenes locally and contexts and histories of organisms would not matter that much when it comes to their particular experiences.

Now comes the really crucial account of supervenience. It is the distinction between logical and natural supervenience. Chalmers needs logical supervenience to run his zombie argument through. With the logical supervenience the notion of supervenience is spelled out in terms of meaning. We can say that *B*-properties supervene logically on *A*-properties if no two logically possible situations are identical with respect to their *A*-properties but distinct with respect to their *B*-properties. It seems rather clear that the logical supervenience connects supervenience and necessity. It is very important to note that the constraints on logical supervenience are mainly conceptual. For example, a world that contains male vixens is conceptually incoherent and therefore impossible, whereas a world containing flying telephones seems possible because there is no obvious incoherence in the idea of flying telephones and given that our world might have turned out otherwise, it seems possible. Given that logical supervenience is defined in terms of logically possible worlds, it makes it completely transparent that when evaluating a statement in terms of logical supervenience a mere lack of contradiction or incoherence would make a statement metaphysically

possible, according to Chalmers. Furthermore, Chalmers claims that if the *B*-properties logically supervene on the *A*-properties, then the *A*-properties entail the *B*-properties. A corollary of this would be that if *A*-facts logically supervene on the *B*-facts, then it should be impossible for the *B*-facts to hold without the *A*-facts holding.

Before proceeding it is worth noting that there could be a supervenience relation without logical supervenience. Chalmers calls it "natural supervenience" or as it is sometimes called in the literature the "nomological supervenience". It is a weaker variety of supervenience that is based on the systematic correlation between two properties in the natural world, that is, a world that does not violate any law of nature. He claims that natural supervenience holds when two situations, that could occur in nature, with a certain distribution of *A* properties have the same distribution of *B* properties. In another words, natural supervenience applies when *A-facts* naturally necessitate *B-facts.* Basically, this variety of supervenience is a relation between two occurring properties according to which, say, *A* properties systematically and lawfully accompany *B* properties. For example, biological facts do supervene on the physical facts naturally, but also logically. It is conceivable and thus possible that the actual world could have turned out differently, but in that case totally different laws of nature would apply and if that would be the case then the dependency would hold as necessary. On the other hand, indiscernible duplicate of our world in respect to all physical properties has to be indiscernible exactly in the same way in respect to the biological properties. Chalmers puts it in yet another way: B-properties supervene naturally on A-properties if any two *naturally possible* situations with indiscernible A-properties have indiscernible B-properties. A naturally possible situation is one that could actually occur in nature, without violating any natural laws. *A* supervenes nomologically on *B*, when for all objects *x* and *y* and all possible worlds *wi* where our laws of nature apply, if *x* at *w1* and *y* at *w2* are indiscernible in their *A*-properties, they are also indiscernible in their *B*-properties.

One can think about natural supervenience in such a way that logical supervenience implies natural supervenience, but not vice versa. I think it is easy to see why that is so. If property *A* is logically supervenient on property *B*, the dependency, although spelled out in terms of meaning, has to maintain a lawful form, otherwise it would be incomprehensible. On the other hand, as we have seen with the example of biological properties and physical

properties, nomological or correlational dependency is not spelled out in terms of meaning and thus does not hold necessary. Chalmers needs natural supervenience in part to describe his ontological position, i.e. property dualism. Now after the requisite notions for understanding Chalmers's account have been explained, let's turn to his critical move in the argument. That is the link between supervenience and necessity and establishing that phenomenal is not logically supervenient on the physical. Here we have a link between supervenience and necessity.

Chalmers holds that higher level properties, or for that matter higher level facts, have to necessary logically supervene on the respective lower level properties or facts for an explanation to be successful. He claims that there are basically three methods for establishing logical necessity, it is through conceivability, epistemology and analysis. More importantly he uses these three approaches in arguing against logical supervenience of the phenomenal on the physical in all five of his arguments. Not all five arguments will be rehearsed here because original versions of some of them are already presented or will be presented in great detail[17]. We just need a basic conceptual framework to understand his account.

The conceivability approach relies on the claim that if the *A* properties logically supervene on the *B* properties at all, then instantiation of *B* properties without instantiation of *A* properties should not be possible.

Epistemologically, if the *A* facts logically supervene on the *B* facts at all, then knowing *B* facts would *a priori* imply knowing *A* facts, at least through the primary intension, Chalmers claims.

Logical supervenience holds *iff* the analysis of the intensions of the *B*-properties in sufficient detail enables one to infer that *B*-statements follow from *A*-statements in virtue of these intensions alone.

[17] I refer here to: Kripke's modal argument, the knowledge argument and two-dimensional argument.

Clearly Chalmers's strategy now becomes much more apparent. Conceivability approach is behind his zombie argument. Epistemological approach supports the knowledge argument and finally method of analysis is employed in the two-dimensional argument.

It is clear by now that entire Chalmers's strategy hinges upon two crucial points: a) the link between conceivability and metaphysical possibility, b) on his account of the phenomenal concept semantics, that is on the reference via descriptions of the causal roles. In the chapters three and four I will be challenging both points b) and c).

Let's consider the inference from logical possibility to metaphysical possibility in his argument. Obviously, mere lack of incoherence or contradiction in conceiving of a situation suffices only for establishing logical possibility. Some further means are required for establishing metaphysical possibility. Let's remind ourselves that logical possibility concerns only the primary intension, that is, *a priori* understanding of how the reference is fixed in the actual world; whereas metaphysical possibility relies on the secondary intension, thus depending on *a posteriori* analysis of counterfactually fixed reference. It is *a posteriori* because it depends on how the reference is fixed according to primary intension, which of course can only be determined *a posteriori*. This is the point where two-dimensional semantics steps in, and it will be thoroughly discussed in the section on two-dimensional argument. Let's just see now how the whole story is integrated here. According to the two-dimensional analysis, in the case of natural kinds like "water" and "H_2O" their secondary intensions are the same although their primary intensions are different. The fact that their primary intensions are different makes it logically possible that "water" is not "H_2O", it is coherently conceivable that something else could have played the role of "water". However, given that according to their secondary intensions both concepts pick out the same thing, it reflects our knowledge about how the reference is fixed according to the primary intension, i.e. it reflects our *a posteriori* knowledge, and makes the situation metaphysically necessary. To be precise, what makes it metaphysically necessary is that when we consider a statement like "water is H_2O" and look at its primary intension we can see that primary intensions of the concept "water" and the concept "H_2O" are different, they both say whatever plays "watery" role is "water" or "H_2O". We say then that it is logically possible that something other than "water" might have played the "watery" role because, after all, the whole world might have turned out differently. However, when we look at their secondary intensions we

can see that both concepts play the same causal role, thus they both pick out the same thing. So when considered through possible worlds the statement "water is H_2O" will be *a posteriori* necessary because it is metaphysically impossible that "water is not H_2O".

Contrary to the case of natural kinds, in the case of zombies no such thing as discarding a counterfactual situation as a misdescription is available, because what feels as pain is pain, and it can not be something else. It is or it is not pain. If it were logically possible that "C-fibers firing" occurs without pain, then it would be metaphysically possible that "C-fibers firing occurs without pain" straight off. The argument goes through in establishing that the phenomenal does not logically supervene on the physical with metaphysical necessity.

Chalmers's departure from Kripke's modal argument is hopefully apparent now. Kripke sets up the argument in terms of essential properties whereas Chalmers bases his arguments on meaning. Of course, meaning based supervenience avoids few of the pitfalls of the Kripkean approach, most notably the objection that essential properties as rigid designators involve arbitrariness to a certain degree. However, as we shall see, meaning based refutation of physicalism that relies on the supervenience characterization actually stand less chances than Kripke's identity based characterization. Naturally, one does not have to accept essential property route in rigidifying phenomenal concepts but instead one could ground the identity claim on meaning as well, as it will be shown in the chapter on phenomenal concept strategy.

Furthermore, Chalmers holds that materialism defined as a psychophysical identity thesis, as it is assumed in Kripke's argument, cannot account for the metaphysical necessity as the supervenience argument does. Chalmers maintains that truly reductive explanation requires a logical supervenience relation, and if that is so, then the materialist must be committed to the thesis according to which phenomenal concepts have to be *a posteriori* and to have necessary primary intension, which means it has to be *a priori*. We will come to this issue later on in the discussion.

However, it is important to note that the supervenience based characterization of physicalism implies that if the phenomenal is supervenient on the physical (regardless of whether it is global or local) it has to be necessarily supervenient, that is, it has to supervene logically. If a situation in which there were physical properties whereas phenomenal

properties were completely lacking is coherently conceivable, then it is logically possible that the phenomenal is not supervenient on the physical. Chalmers holds that if such a situation is logically possible, then it is metaphysically possible, thus making physicalism necessary false. The step from logical possibility to metaphysical possibility here might seem dubious, but Chalmers fortifies it with the two-dimensional semantics. We will come to this issue shortly. The main reason why the zombie argument goes through in the case of phenomenal properties and breaks down in the case of natural kinds, according to Chalmers, is because of the assumption that conceptual semantics is based on the descriptions of causal roles of the properties they refer to. So whereas in the case of natural kinds one can admit that it is logically possible that "water" is not "H_2O" but something else, say "XYZ", because the actual world might have turned out differently, the counterfactual situation in which "water" were not "H_2O" is discarded as a misdescription, thus making it metaphysically impossible that "water" is something else than "H2O". Let me clarify this. One can discard a counterfactual situation in which "water" were not "H_2O" as a misdecription because we get to know what plays the causal role of "water" through the third person perspective, because for all we know something else indeed could play that role. On the other hand, what it is like to be in pain, is characterized by that very feeling of what it is actually like to be in that state. In other words, we get to know about these properties only through the first person perspective. If something looks like pain and feels like pain then it must be pain. It is due to this feature of phenomenal knowledge that we cannot discard the logical possibility of zombies as a misdescription, and it is because of this that in the case of the phenomenal the argument goes through to metaphysical possibility. As we shall see in the third chapter, that doesn't have to be the case with phenomenal concepts. More on that will be said then.

Let's just briefly review the argument from supervenience as it will be discussed in a bit more details in the chapter on the phenomenal concept strategy.

1.5.3 The supervenience argument

Let's remind ourselves on the formal supervenience argument from the beginning of this section.

20. The fact that conscious experience exists is a positive fact;

21. Conscious experience is not logically supervenient on the physical;

22. If there are positive facts that are not logically supervenient on the physical facts, then physicalism is false;

---

23. Therefore, physicalism is false and phenomenal properties are further properties of the world.

Once premise 21. is established in the zombie argument, Chalmers has an open road to cast the argument from supervenience. The first premise is granted by the mere immediacy of the first person experience. Chalmers even says that we could doubt about the existence of other minds, but one thing that is for sure is the unmediated intimate acquaintance with our own experience.

Again, as with the zombie argument, some of the premises in the argument can be challenged, and have been challenged, but the conclusion cannot be questioned if one accepts all the premises. Clearly, denying the first premise commits one to eliminativism. We won't be spending much time here on this approach as it was already outlined in the previous subsection.

My route shall be as in the case of zombie argument to argue against premise 21. The approach that I would like to pursue here shall be twofold. On the one hand, I will be arguing from the perspective of the phenomenal concept strategy a la Papineau, and, on the other hand, I will be arguing that arguments for both ontological and explanatory gap positions in fact exemplify a case of brute disagreement no argument could help overcome. It rather

seems that the dispute is about the clash of intuitions and that only an account of how we connect psychological concepts and phenomenal concepts can help us understand the perplexity of the problem. But before I can begin laying out my key argument some more exegetical work needs to be done.

In order to understand the technical subtleties of the metaphysical and epistemological assumptions involved in the zombie/supervenience argument we will need to rehearse and analyze the two-dimensional argument. Let's turn now to this issue.

1.6 The 2-D argument

As we probably know by now, the conceivability argument starts with the thesis to which a materialist must be committed: if the mental states supervene on the brain states ($P \Rightarrow Q$), then a situation in which there are brain states without mental states should not be possible. This is so because, according to Kripke (Kripke 1980), all identity statements involving rigid designators, if true at all, must be necessary true. After having established what a materialist is committed to, the argument departs from the claim that a situation in which the brain states take place but phenomenal states are lacking altogether is conceivable. It is argued, then, if such a situation is conceivable, it is possible. Finally, if such a situation is possible, then materialism is false, because as we recall identity has to hold necessarily in order to be true.

Now, how do we come from a purely epistemic claim about what is conceivable to the modal claim about what is possible?

According to Kripke, the only thing we need to make this step is a lack of contradiction in conceiving such a situation. This principle has often been criticized in the case of zombies. Now the famous example with the natural kind terms comes into play. Namely, there is no contradiction in conceiving a situation in which water is not H_2O either. However, because "water" rigidly designates H_2O, the argument actually shows that, although conceivable, it is not possible that water is not H_2O. How does this step of the inference work? Since it is an empirical discovery that water is H_2O, the reference of the

identity statement is not fixed *a priori*. In other words, the state of affairs in our world might have been different, so water might have turned out to be something else in our world (depending on the actual empirical discovery). Therefore, there is nothing in our knowledge that would contradict the conceivability of the claim "water is something other than H_2O". Just to give a contrastive example, given the definition of a triangle, a geometrical figure that has three angles, it is neither *a priori* conceivable nor possible for a triangle to have four angles.

Why does this step from conceivability to metaphysical possibility break down in the case of "water⇒H_2O"? According to Kripke, it is because the reference of the concept of water is fixed to its microphysical structure, which in our world is H_2O. On the other hand, if it turned out that the microphysical structure of water in our actual world was not H_2O but something else, say XYZ, then the reference of "water" would be fixed to XYZ in the actual world. According to Kripke, rigid designation of natural kind terms is always tied to microphysical structure. This, on the other hand, means that whatever we discover at the microphysical level to play the role of water, fixes the reference of "water". In a sense, it is not the empirical discovery that fixes the reference, because the reference is fixed beforehand, but whatever we believe to play the role of water. This is how rigid designation of the natural kind terms actually works according to Kripke. Therefore, it is only conceivable that water might not have been H_2O, but it is not possible. On the other hand, if some counterfactual world at which water is not H_2O, was actual, then "water is H_2O" would be false considered from that world. And this is where the two-dimensional semantics steps in. According to Chalmers's interpretation of two-dimensionalism, concepts have two intensions, which are functions from truth conditions to possible worlds and vice versa. The intension that goes from the truth conditions in the actual world to the counterfactual worlds is *the primary intention*. The primary intension tells us what fixes the reference in the actual world. The secondary intension goes from possible worlds to truth conditions in the actual world and tells us how the reference in the counterfactual world is fixed. In other words, a term S is verified in some possible world W considered as actual, if and only if its primary intension is true at W. A term S is satisfied in some possible world if and only if its secondary intension is true there. The two-dimensional semantics is here employed to show why the step from conceivability to metaphysical possibility in the case of

phenomenal/physical works and why it breaks down in the case of water/H_2O. Let's have a closer look at what is going on here.

In order to utilize his interpretation of two-dimensional semantics, Chalmers distinguishes several kinds of conceivability. Conceivability is here to be understood as an epistemic notion. The first one is what he calls "*prima facie* conceivability", which requires that the subject should not be able to rule out a hypothesis solely by *a priori* reasoning. The second one is ideal conceivability which goes a bit further in trying to abstract from the cognitive capacities of a subject, and claims that a hypothesis is ideally conceivable if it cannot be ruled out *a priori* even on ideal rational reflection (Chalmers 2009).

The opposite case from these two is the already mentioned example with a triangle - given the definition of triangles as geometrical bodies with three angles, conceivability of the triangles with four angles is ruled out *a priori*.

These two kinds of conceivability are examples of negative conceivability. However, there is also a positive conceivability. This sort is best known from Descartes's thought experiments (Descartes 1996). Positive *prima facie* conceivability is the case when a subject can imagine a situation that she takes to be coherent and as the one in which the hypothesis is the case. As with negative conceivability, here we also have ideal positive conceivability. We say that a situation is ideally positively conceivable if its *prima facie* positive conceivability cannot be defeated.

However, the most important kinds of conceivability according to Chalmers are primary and secondary conceivability. These two kinds of conceivability can be based on either positive or negative conceivability. What distinguishes them is not just an epistemic perspective or cognitive capacities, but also a sense in which we say something is conceivable. So, we say, for example, that "water is not H_2O" is primary conceivable, exactly as it is stated, that is, that water is actually NOT H_2O (which in Kripkean terms means that the whole world is otherwise, not just water and H_2O), whereas in the case of secondary possibility we take a sense of conceivable in which "water is not H_2O" merely seems conceivable but is not conceivable. An example for the sense in which we think of the secondary conceivability is to say that when we say "water is not H_2O" we only mean that watery stuff (that seems like water) is not H_2O, but water as we know it is still H_2O.

When we look at the distinction between primary and secondary conceivability in the light of the distinction between the *a priori* and the *a posteriori* knowledge, it is clear that primary conceivability can only be made plausible in the domain of the *a priori*. That is, a hypothesis is primary conceivable if it cannot be ruled out purely on *a priori* reasoning, without any reference to empirical or extra-linguistic knowledge.

So what is the link between primary and secondary conceivability on the one side and metaphysical possibility on the other?

The primary intension of the sentence "water is not H_2O" (¬S hereafter) is true at the counterfactual world W considered as actual, and its secondary intension at W1, say our world considered as counterfactual, is false, because at our world water is H_2O. To put it differently, a world verifies S iff its primary intension is true at that world, and a world satisfies S iff its secondary intension is true there considered from the actual world.

So, in order to prove that these steps of inference in the conceivability argument follow *a priori*, Chalmers's reasoning is based on the distinction between primary conceivability and possibility and secondary conceivability and possibility. So how does this work? We said that, for example, water is not H_2O is primary conceivable and therefore primary possible because "water is H_2O" is not an *a priori* truth. Knowledge of "water is H_2O" is not based on the meaning of the terms, linguistic rules, etc, but it is an empirical discovery, which we then employ to understand meaning of the terms. So, for all we know water might have been something other than H_2O in our world. Unlike the case of a triangle, wherein a four-angled triangle is simply incoherent even to conceive of, because if we grasped the meaning of the concept "triangle", we know, without any reference to extra-linguistic or to empirical knowledge, that a four angle triangle is an incoherent concept, and therefore not possible. So when we say "water is not H_2O" is possible, it only means that there is nothing in our knowledge that would contradict to this statement *prima facie*, in a sense, this is not a statement about just water and H_2O, but the statement about the whole world and its nomological structure. Now, Chalmers wants to show that if we apply the conceivability argument to the case of water/H_2O, what is actually claimed simply leads to inconsistency, by showing that although in the case of water/H_2O the statement "water is not H_2O" is *prima facie* conceivable, thus primary possible, it is not conceivable and therefore not possible in

the sense in which water is still H_2O and there is some watery stuff that has the same macro properties of water, but which is not H_2O, that is "water is not H2O" is not secondary possible. In the case of primary possibility, if water was not H_2O but otherwise, then as Kripke puts it, the whole world would be otherwise and the statement "water is not H_2O" would be true. This is why two-dimensional analysis comes very handy in these complicated cases. We need the secondary possibility to see what is going on with the secondary intension considered from some centred world.

Let us the structure of the argument with natural kinds like water and H_2O. Let *W* be the proposition that water is not H_2O.

24. *W* is conceivable;

This premise says that *prima facie* (it could not be ruled out *a priori*) it is conceivable that water is not H_2O.

25. If *W* is conceivable, *W* is 1- possible;

This premise says that if the statement is *prima facie* conceivable, then it is primary possible.

26. If *W* is 1- possible, then *W* is 2-possible;

This premise says that if the proposition "water is not H_2O" were primary possible, then it would be secondary possible, which doesn't follow in this case.

Whereas in the case of the phenomenal/neurobiological the argument goes right through, from primary possibility to secondary possibility, and given the physicalist commitment thesis, the conclusion is not compatible with it - therefore physicalism is false. Let us have that argument formalized as well:

27. (P&¬Q) is conceivable;

28. If (P&¬Q) is conceivable, (P&¬Q) is 1- possible;

29. If (P&¬Q) is 1- possible, then (P&¬Q) is 2-possible;

30. If (P&¬Q) is 2-possible, materialism is false;

31. Therefore, materialism is false.

So what is going on here compared to the water/H_2O case? The argument here says that if a hypothesis is conceivable then it is primary possible, in another words it is not possible to rule out that hypothesis solely on the ground of *a priori* reasoning. This is the point where epistemic and modal principles meet, because if a conceived hypothesis is not possible to rule out *a priori*, then it is possible. This is the meaning of primary possibility, or the link between conceivability and possibility in Kripkean terms, which has been criticized in particular for this step. In an attempt to block these principle objection, Chalmers goes further in claiming that although the statement "water is not H_2O" is primary possible, the identity of water with H_2O is established only *a posteriori*; however, once established as an identity it holds necessary. We might say that there are no *a priori* reasons that would contradict to conceiving of a situation in which water is not H_2O in a strict sense (something other than H_2O plays the water role) - this is a sense of the primary possibility – but it is not possible that water is H_2O in the actual world and yet something else in a counterfactual world - this is the sense of the secondary possibility. This distinction clearly shows that primary conceivability does not lead to secondary possibility in the case of "water is not H_2O". However, "water is not H_2O" is compatible with the inferences from primary conceivability to primary possibility and from secondary conceivability to secondary possibility.

The argument against the identity of water and H_2O breaks down when it comes to the inference from primary conceivability to secondary possibility, because the statement has the primary and secondary intensions that do not coincide. In other words it is not coherent to think of the situation in which a world W1 verifies the statement's primary intension and a world W2 satisfies its secondary intension, given that they are separated (do not coincide), that is it is not coherent to hold that water at W1 is H_2O and water at W2 is XYZ, since "water is H_2O" is verified at W1 (is necessarily true) it cannot be XYZ. On the other hand, in the case of phenomenal/neurobiological the primary and secondary intensions of the statement do coincide, which in effect allows for the inference from primary possibility to secondary conceivability, and further to secondary possibility. This is so because, as Kripke noted, in the case of phenomenal/neurobiological there is no strong dissociation between appearance and reality, to paraphrase him: pain is identified with its immediate quality. To

put it into more technical terms, dissociations between water and H_2O, or between heat and molecular motion, seem conceivable, but they are only primary conceivable, thus primary possible, because their meanings have two modal dimensions, one that goes from the possible world to truth values or primary intension, which tells us what plays the water role in the centred world; and the other intension that goes from truth values to possible worlds or the secondary intension. Now, since in the case of natural kind terms, primary and secondary intensions do not coincide, it is because of the way in which we learn about their primary intensions (objective, third person perspective), in the case of phenomenal consciousness primary and secondary intensions collapse into one, because of the immediacy of phenomenal knowledge which dissolves the distinction between appearance and reality, in a sense that whatever properties of our experience appear to us, they are not merely appearing, they are real. Remember Nagel's account (Nagel 1974).

What is going on in the case of phenomenal/neurobiological is that it seems that both W1 and W2 verify the statement. In a sense primary and secondary intensions of the statement "the phenomenal is not neurobiological" seem to coincide. That is the reason why the two-dimensional argument in this case goes through from primary possibility to secondary conceivability and further to secondary possibility. This argument structure purports to be saying that both statements - "pain=C fiber firing" and "C fiber firing≠pain"- seem to be true, which is a contradiction.

To put it in other words, in the case of natural kind terms, like "water" or "heat", there seem to be a clear dissociation between what is just an appearance and what is real, i.e. we can imagine that something might look like water or heat whereas in fact it is not. Since the identity statement "water is H_2O" is not based on *a priori* knowledge, there is no *a priori* contradiction in conceiving of a situation in which water is not H_2O, and thus if the hypothesis cannot be ruled out *a priori*, then it is possible. What makes it break down is the fact that given the *a posteriori* necessity of the identity statement "water is H_2O", it is not possible that water is something else at W2 and still H_2O at W1. However, what is possible is that if water was something else in the actual (centred) world, then its primary intension would be fixed to whatever that "something else" is, to whatever plays the water role in the counterfactual scenario. In this case the statement "water is not H_2O" would be true in the actual world. But this only means that the whole world would be otherwise. This makes a

strong case, then, that primary and secondary intensions of phenomenal terms do in fact coincide. This in effect makes the two-dimensional argument involving the phenomenal/neurobiological to go through to the secondary possibility, which ultimately means that physicalism leads to contradiction, assuming that the primary and the secondary intensions of phenomenal terms coincide.

In this chapter we started off with Nagel's account of the epistemic gap and proto-knowledge argument (Nagel 1974). We examined issues of subjective and objective knowledge and its role in determining modal and ontological status of a statement (Kripke 1980). We also saw how the epistemic gap according to some philosophers (Jackson 1982, 1986) generates very serious metaphysical problems. We saw that these accounts need to be backed up by considerations about the inference from conceivability to metaphysical possibility and about two-dimensional modal semantics (Chalmers 1996, 2009). In the next chapter I will be considering in great detail Levine's account of the explanatory gap (Levine 1983, 1993, 2001) that utilizes the distinctions and arguments presented in this chapter. After that I will be presenting and discussing my own account of the problem, starting from chapter three until the very end of the dissertation.

2. CHAPTER 2

2.1 The explanatory gap account

"On the one hand, we have excellent reason for thinking that conscious experience must be reducible, in the requisite sense, to a physical phenomenon, and, on the other hand, we don't see how it could be." (Levine 2001, p. 175).

As we have seen in the previous chapter, the anti-physicalist conceivability arguments force us to face serious metaphysical consequences if we accept all their premises. Nagel's account (Nagel 1974) was just an outline of the knowledge argument (Jackson 1982, 1986). Jackson argued further that from the body of physical facts one cannot infer phenomenal facts, and that, therefore, phenomenal facts are not physical facts. Both of these two accounts presuppose that the conceptualizations available for the neurophysiological and the phenomenal domains respectively are at the very least incompatible. This roughly means that concepts that we use to grasp the neurophysiological properties leave out some of the crucial or defining features of the phenomenal properties. It is the "what is it likeness" of experience from the experiencer's point of view. In the first chapter, we had an opportunity to follow this line of reasoning in an ascending order of complexity from Nagel (Nagel 1974), Kripke's identity based modal argument (Kripke 1980), Jackson's knowledge argument (Jackson 1982, 1986), Chalmers's zombie argument and the argument for property dualism (Chalmers 1996) to the two-dimensional argument (Chalmers 2009). Now the time has come to examine some of the central assumptions in all these arguments, namely the incompatibility of phenomenal and physical concepts, and the consequences of these arguments. At its core the problem we are about to tackle is that "…we have excellent reasons for thinking that mental phenomena, including conscious

experience, must be a species of physical/natural phenomena. On the other hand, we also have excellent reasons for thinking conscious experience cannot be captured in physical/natural terms" (Levine 2001, pp. 9–10). Levine powerfully argues that although the entailment from conceivability to metaphysical possibility of zombies cannot be established, materialism is still not off the hook because this claim needs a further argument that would explain how the same property can instantiate two semantically different concepts. In this chapter I lay out his account along with the technical/conceptual apparatus it rests upon. This whole chapter is devoted solely to laying out the explanatory gap account and therefore it is shorter than the other chapters. At the end of the chapter I will re-examine the whole account and provide a criticism of it. After that in the third chapter I pave the way to my own position by assessing some of the most powerful responses to the explanatory gap.

Let's turn now to the explanatory gap account. I start off laying it out by examining some technical terms Levine (2001) has introduced. The first on our agenda is the difference between "thick" and "thin" modes of presentation and "thick" and "thin" concepts. After that I will be examining notions of "thick" and "thin" conceivability, ascriptive and non-ascriptive modes of presentation, then gappy and non-gappy identities, and finally in the last section of this chapter I will be concerned with the issue of the intelligibility of explanations. Let's begin with "thick" and "thin" modes of presentation.

2.2 Thick and thin concepts and modes of presentations

To understand this idea one must realize that the ways in which we conceptually grasp properties determine the way in which the explanation of a given phenomenon works. One can think of the ways in which we think about content of our experience as modes of presentation (Levine 2001, p. 8). Levine claims that this point becomes really obvious when one considers the modes of presentation that are employed when we think about natural kinds and about qualities of our own experiences. In the case of natural kinds the mode of presentation is representationally "thin", i.e. our cognitive relation to the content of experience of a natural kind bears nothing substantial or determinate apart from the informational presentation and perhaps relation to other presentationally "thin" properties.

Take for example a concept of a cat. There are two things that are relevant for the mode of presentation here. On the one hand, there is a mental symbol “cat” and on the other hand, there is some kind of a nomic relation between that symbol and a property of being a cat. Obviously, the symbol “cat” plays some cognitive role, but that is not determinate for the content of a concept of cat. What is more important for the content of a natural kind concept is its relation to “cathood”. In this sense, there is nothing substantial or determinate about our cognitive relation with natural kinds concepts and that is why they are presentationally “thin”.

Contrary to this, our cognitive relation to the content of phenomenal experience is substantive and determinate, so in this sense it is presentationally “thick”. In a way, the experience itself serves as its own mode of presentation in this case, for if it were not present the relevant conception could not be formed. This is what is meant by the term “substantive”. To say that a mode of presentation is determinate, means that it presents certain and specific quality in its own right. In the case of “thick” modes of presentation cognitive significance is more relevant for the content of a concept than the information about the relation of a property it represents with other properties. Having distinguished “thin” and “thick” modes of presentation we can say now that phenomenal concepts are “thick” concepts because they rely on the “thick” modes of presentation. On the other hand, natural kind concepts are “thin” concepts because they are based on the “thin” modes of presentation. That is, our cognitive access to the representation of a situation is not in any significant way tied to the experience of that situation. On the other hand, in the case of psychophysical identity our cognitive access to the phenomenal side plays a significant part of the concept of experience. In this case we have a “thick” concept.

The difference between thick and thin concepts in Levine comes down to this: the semantics of the thick concepts is determined by modes of presentation wherein a state or a quality presents itself regardless of the nomic relations a given property might have with other properties. In other words, one could have a thick concept of experience without knowing anything about the nomic or causal relations of the property it picks out with other properties. On the other hand, one could have a thin concept without having the determinate and substantial cognitive access to the content of a thin concept. This situation plays an essential role in all arguments and accounts presented so far.

What is meant by the claim that the physicalist explanation leaves out something is that the two ways in which thin and thick concepts are determined seem incompatible. Natural kind concepts are based on the descriptions of the causal or nomic roles their referents play, whereas phenomenal concepts are based on the immediate and determinate quality of experience that in a sense presents itself.

Why should this be a problem?

Levine holds that the problem lies in the fact that one of the plausible explanations could be that these two ways in which we pick out properties reflect the metaphysical gap between the phenomenal and the physical properties. On the other hand, given that we have very good reasons to believe that materialism is true based on the causal argument, the incompatibility of thick and thin concepts render the psychophysical identity unintelligible even if it were true. We will come to this point very soon. Let's continue with laying out Levine's account.

2.3 Ascriptive and non-ascriptive modes of presentation and apriority and necessity

Our next issues are the modal and epistemological aspects of the explanatory gap problem. Levine's idea about the relation between these two notions can best be articulated through his analysis of the conceivability arguments.

Starting off with the terminological clarifications about what is meant by "metaphysically possible" or "logically possible", Levine (Levine 2001, p. 40) argues that there is no real distinction between them. For him the logically possible is metaphysically possible, or just plainly possible. Instead of that he introduces the distinction between logically/metaphysically possible and conceptually possible. The crucial distinction here is between "situations" and our modal judgments about them. The modal status of a "situation" is mind independent, that is to say they could be metaphysically possible without us knowing about it. On the other hand, when a "situation's" modal status is dependent on our judgements then it is conceptually possible. This point will come very handy later on in presenting Levine's account of what follows from the conceivability arguments. Suffice it say

that it obviously follows from these considerations that a "situation" cannot be metaphysically necessary without being conceptually necessary. Before laying the whole account out, we must explore what Levine takes the connection between the epistemological and metaphysical sides of modality to be.

These distinctions are based on the dependency of our modal judgements upon our epistemic situation. For example, someone could be in a position to judge it as necessary that H_2O contains hydrogen and that water contains hydrogen. But someone else who is ignorant of the chemical composition of water would not judge it as necessary, because that person would not know the relevant description of the causal roles under which both "water" and "H_2O" pick out the same thing, that is she would be unaware of the bridging claim that "water=H_2O". That is to say, our judgements about factual and modal descriptions of the situations are relative to the representations that we use to pick those situations out. Naturally, a "situation" in this context is taken to mean "an object instantiating one or more properties or an ordered n-tuple of objects that serve as the truth conditions for a statement" (Levine 2001, p. 40), that is, a "situation" allows us to evaluate a statement as true or false. Now comes an important part of the argumentation. It is the relation between conceptual possibility and metaphysical possibility. A situation *S*, is conceptually possible, iff when thought of under the representation *R*, it is possible. On the other hand, a situation *S'* is metaphysically possible iff *S'* is conceptually possible under representation *R* and it is not *a priori* that not-*R* (Levine 2001, p. 40). A simple principle then follows, if there is a description of a situation according to which the situation is conceptually impossible, then it is metaphysically impossible (Levine 2001, p. 46).

Let me clarify a very important point about a priority and necessity in this context. Consider a statement "Water contains hydrogen". To render this statement necessary, three crucial requirements must be met. First, that the statement "H_2O contains hydrogen" is *a priori*, i.e. that its truth or falsity can be evaluated solely on the ground of the meaning of the terms involved. Second, that there is an empirically discoverable fact, thus *a posteriori*, that water is H_2O. Finally, the third requirement is that the statement picks out a situation in which water contains H_2O. Based on these preliminaries we evaluate a statement as necessary even if it is *a posteriori*. That is, to render an *a posteriori* statement necessary,

there has to be an *a priori* description from which that statement can be derived. This is so at least in the case of natural kinds.

In a sense each statement that is metaphysically necessary must be also conceptually necessary, and if a statement is conceptually impossible it is metaphysically impossible. Thus far this seems obvious. On the other hand, a statement that is metaphysically impossible is not necessarily conceptually impossible as we have seen in the case of water/H_2O. That is to say, if a situation is conceptually possible that fact alone does not justify entailment to metaphysical possibility. Basically, this point follows from the claim that the only way we can evaluate a situation's modal status is through a representation of the given situation. In Levine's view (Levine 2001, p. 46), the fact that conceptual possibility does not entail metaphysical possibility is exemplified by all cases of *a posteriori* necessity. If physicalism were true it should also fall in this category.

So how does this account relate to conceptual and metaphysical necessity? Basically, this account comes down to two-dimensional semantics. What is conceptually possible is in Chalmers's lingo 1-possible and what is metaphysically possible is 2-possible with Chalmers. What happens in Levine's analysis of the conceivability arguments is not much different from the two-dimensional analysis, at least in terms of basic conceptual assumptions. The point where Levine departs from the standard interpretation of the conceivability arguments is with the epistemic assumptions. Namely, Levine claims that unlike the identities of natural kinds, wherein our epistemic situation is "thin", in the case of phenomenal concepts our epistemic situation is "thick" which makes the psychophysical identity a gappy identity. I will explain what a "gappy" identity is in a moment. Before that I just need to crudely elaborate on Levine's departure from the standard interpretation of the conceivability arguments. Because if the "thick" concepts are not based on the modes of presentation that involve nomic or causal relations between properties, then it is possible in principle that a characterization of a counterfactual situation containing "thick" concepts might be missing some relevant description, i.e. a description that could render the counterfactual situation metaphysically impossible. This in fact leads only to the conceptual possibility of zombies. Because the "thick" concepts are flanking on the one side of the psychophysical identity, that kind of identity is "gappy". It is very interesting to note that the very same assumption about the phenomenal concepts that supports the metaphysical and anti-materialist

conclusions of the conceivability arguments is employed here to block the inference to metaphysical possibility. We will see in the next chapter how the notion of "thick" phenomenal concepts is employed in some of the most influential and successful responses to the epistemic and the ontological gaps, namely in the phenomenal concept strategy.

Another way of explicating this view is through the analysis of the ways in which the conceptual content can be determined. Namely, Levine (Levine 2001, p. 54) distinguishes between ascriptive and non-ascriptive modes of presentation. Ascriptive modes of presentation are those we are already familiar with from Kripke and Chalmers's versions of the conceivability argument. According to ascriptivists the concepts pick out their referents via the descriptions of their standard causal or functional roles. To put it into more technical terms, a concept's primary intension bears almost no cognitive significance; it is almost entirely based on the symbols and their relations. The point that clearly distinguishes between ascriptivists and non-ascriptivists is that whereas ascriptivist believes that the *a priori* knowledge is a matter of having enough information about the causal or functional roles properties normally play, a non-ascriptivist maintains that *a priori* knowledge cannot be founded in such a way.

On the other hand, the non-ascriptive mode of presentation works in such a way that the content of a concept is determined by the formal relations a given property has with other properties.

2.4 Thick and thin conceivability

Let's turn our attention now to rather technical aspects of the explanatory gap account. The claim that the conceivability argument only establishes an epistemic possibility of counterfactual scenarios has been a subject of many vigorous debates over the years[18]. For the purpose of current work and sake of brevity we won't go into details of these

[18] Cf. (Chalmers & Jackson 2001), (Block & Stalnaker 1999), (Stoljar 2000, 2001), (Levine 1998, 2001), (Balog 1999), (Hill 1997), etc.

debates. However the argument that separates the epistemic possibility from the metaphysical one needs to be rehearsed.

This issue is of central importance to the explanatory gap argument and now we will consider it in some detail. We have seen in the contrasting cases of natural kind identities that the conceivability argument breaks down when it comes to the entailment from conceivability to possibility and that it goes right through in the qualia case. According to Levine's interpretation of the argument, it is so because metaphysical possibility relies upon conceptual possibility. Namely, we get to the metaphysical possibility not directly from conceivability but through conceptual possibility first. The notion of conceptual possibility only requires that a counterfactual situation that we are conceiving of should not be inconsistently inferable from the situation we start with. If a situation is conceptually possible, it is metaphysically possible as well (Levine 1998, p. 454). Since there is no inconsistency in conceiving of a situation "brain state *B* but not quale *M*" it is then conceptually possible that "brain state *B* but not quale *M*", and if this is conceptually possible then it is metaphysically possible. This is the standard "metaphysical reading" of the explanatory gap according to Levine. What he needs now is a move that undermines the inferential link between conceptual and metaphysical possibility, thus establishing only epistemic possibility of counterfactual cases of qualia. Levine argues that there could be two ways to understand the inferential link, by analyzing the ways concepts connect to their referents, or the concept's modes of presentation (Levine 1998, pp. 457-8). On the one hand they could involve ascriptive modes of presentation and on the other non-ascriptive modes of presentation (Levine 1998, 2001). Ascriptive modes of presentation are, normally, involved in standard cases of natural kind identity, wherein we have two distinct concepts that pick out the same thing. This case involves ascription of properties to the referent by way of instantiating the properties that are referred to; for example, the causal roles of water. The non-ascriptive mode of presentation uses other methods. The crucial difference between the two is that in the case of ascriptive mode the properties are only contingently related to each other, and perhaps even cognitively represented. Whereas in the case of the non-ascriptive mode properties are related in some other way and not necessarily cognitively represented. According to Levine, we only need non-ascriptive modes of presentation to undermine the link between conceptual possibility and metaphysical possibility in the case of zombies (Levine 1998, 2001), because if they are non-ascriptive

then we cannot know whether something is missing from the zombie description and therefore allowing for the inconsistent description of zombies. In other words, non-ascriptive modes of presentation undermine the inference from conceptual possibility to metaphysical possibility.

To see more clearly how this blocking works let's consider it in terms of conceivability. Levine distinguishes two grades of conceivability: thin and thick conceivability (Levine 1998, p. 468; 2001). A thinly conceivable situation is based on a plain conceptual possibility, i.e. a situation is thinly conceivable *iff* it is not inconsistent to infer it from the actual situation. On the other hand, a situation is thickly conceivable if it is conceptually possible and if any inference from its conceptual possibility to metaphysical possibility is based on gappy identities. In this way the link between conceivability and metaphysical possibility is undermined.

This view is unfortunate for the materialists because it implies that the only way to defend materialism is by conceding on the gappy psychophysical identity. On the other hand, to say that a situation is epistemically possible is to say that there is nothing in our knowledge that would make it metaphysically impossible.

2.5 Gappy and non-gappy identities

The point of gappy and non-gappy identities comes down to this: gappy identities require some explanation of how these two concepts pick out the same property. Non-gappy identities imply distinct properties. If zombies are conceivable then it is so because either psychophysical identity is a gappy identity or there is no identity after all. Now how does this stop the flow from conceptual possibility to metaphysical possibility? In the natural kinds cases a situation can be conceptually possible but metaphysically impossible nevertheless. If we want to break the inference from conceptual possibility to metaphysical possibility in the case of qualia, we need to use gappy identities and claim that proper identities need no explanation of their truthfulness. However, it is because of this manoeuvre that we need a further explanation of how the two distinct concepts pick out the same property. This is what is meant by the claim that EG saves materialism from the metaphysical consequences of conceivability arguments but at the price of the explanatory gap.

Levine claims that there are two main aspects of the EG. One involving incompatibility of thick and thin concepts and the other involving gappiness of the psychophysical identity, which is based on the intelligibility of the further request for an explanation. My idea is that gappiness in Levine's case is actually based on the incompatibility of thick and thin concepts which in effect dissolves very easily when treated with Papineau-style phenomenal concept semantics.

2.6 Intelligibility of explanations

The explanatory gap argument is supposed to show that the metaphysical conclusions standardly drawn from the conceivability arguments do not hold but at the price of the explanatory gap. To elaborate on this point Levine (Levine 2001) distinguishes between two aspects of an explanation: the metaphysical explanation, according to which we explain property A in virtue of property B, and an epistemological explanation, which is supposed to explain why property A is explicable in virtue of property B. In another words, epistemological explanation makes it intelligible why given B there is A. They are of course interconnected and interdependent, but it is crucial to note that they are not the same thing. For example, if one were to try explaining my visual experience of a red tomato metaphysically, the explanation would appeal to the fact that at certain point in time *t,* when I had an experience *S*, my brain was occupying a state *B* at *t.* However, to make this explanation intelligible we need a theory of realization that would tell us how *B* realizes *A,* according to Levine.

There is another aspect of the explanatory gap, which concerns mainly the issue of explanations and identities. Take, for example, our contrasting cases of natural kinds and phenomenal kinds. Levine (Levine 2001, 2006) agrees that proper identities need no additional or further explanations. If being in a brain state *B* were the same as having a "reddish" quale *Q*, then it doesn't make much sense to ask what makes this identity statement true. However, it is legitimate to ask why given brain state *B* there is the quale *Q*. In the case of natural kinds an answer to this question can take only two routes according to Levine (Levine 2001, p. 81). One way to explain it is by relying on the basic or fundamental law; the other way is through the identity premise itself.

Let's turn our attention now to the issue of explanation in the explanatory gap and try to connect the notions elaborated above.

Levine (Levine 1983, p. 357) argues that even *if* we succeeded in explaining the causal or functional roles of qualia, there is still the qualitative character of our sensory experiences that is left out by this sort of explanation. For a property to be left out from a theory mean that the theory cannot refer to these properties (Levine 1993, p. 121). Unlike cases of identities of natural kinds, like water and H_2O or heat and molecular motions, wherein a requirement for further explanation does not seem intelligible, because the seeming contingency can be explained away by either deriving the identity from more fundamental statements or dismissed as a misdescription, in the case of qualia this requirement does make very much sense. Levine (1999) holds that most of the natural kind identities are not self-evidently explicable either, and that we can legitimately ask for further explanation. However, after providing the descriptions from which the natural kinds flanking the identity statement are derived, we come to realize that the request for further explanation is just a request for a justification. On the other hand, in the qualia case no derivation seems possible and the puzzlement about how qualia and brain processes could be identical deepens even more as we add refinement to our concepts of qualia or of brain processes. Levine holds that it is so because the psychophysical identity is a "*gappy identity*" (Levine 1998, 1999, 2001). A gappy identity is an identity statement which makes it intelligible to request for further explanation. Of course, non-gappy identity is the one for which it is not intelligible to require further explanation. To illustrate this point let's compare two cases of identity. One involving the boiling point of water and the other involving the presence of the reddish quale. Let's start with the boiling point of water:

32. "H_2O molecules exert vapor pressure *P* at kinetic energy *E*;

33. At sea level exerting vapor pressure *P* causes molecules to rapidly escape into air;

34. Rapidly escaping into air is boiling;

35. 100°C is kinetic energy *E;*

36. Water is H_2O;

37. Water boils at 100°C at sea level." (Levine 1999, p.4)

This is clearly the case of a non-gappy identity according to Levine. One can legitimately ask why 36. is true, but once we provide enough details of the causal roles of water, or we fill in the microphysical description of the behaviour of water, it seizes to make sense to ask for further explanation. In a sense, scheme of the causal roles of water and sufficiently refined microphysical description exhaust all there is about an explanation of the natural kinds identity. We are not left with a sense that something is missing. It is, however, crucial to note that in this example 36. serves as a binding point for the reference of water and H_20 in the explanation. It identifies the phenomenon to be explained with a phenomenon that is describable in terms of microphysics. It enables 36. to be derivable from the combination of statements that are either *a priori* or are descriptions of the underlying microphysical phenomena, which in effect makes it unintelligible to ask for further explanation of 36.'s truth. Furthermore, if we try to abstract the formal structure of the derivation, it comes down to this:

38. "Water is the stuff that manifests the "*watery*" properties;

39. H_2O manifests "*watery*" properties;

40. Water is H_2O." (Levine 1999, p. 5)

It is important to note that premise 38. is analytic and that the characterization of properties in question is given in *"topic neutral*" terms. To say that premise 38. is analytic in this context means that one can know it is true solely on the ground of knowing what the terms involved mean. "Topic neutral" here means simply that the terms that pick out properties in question do not contain any vocabulary that is not contained in the theory that does the explaining. These two features of the binding premise are crucial for the derivation, because ultimately, the explaining of the identity must end up either with the fundamental law or it has to be derived for the identity premise itself, otherwise it would not be an explanation at all. Let's now consider the case of reddish quale.

41. "*S* occupies brain state *B*;

42. Occupying brain state *B* is to experience a reddish quale;

43. *S* is experiencing a reddish quale.“ (Ibid)

In this case we have a gappy identity. Regardless of the amount of detail that we could stack in between 41. and 42. it still makes sense to ask why is 42. true. It is because of the nature of the binding premise 42. As we have seen, premise 38. enables the macro phenomenon to refer to the microphysical descriptions or to the combination of analytic statements. Premise 42., however, considered as a binding premise, the premise that is supposed to identify the phenomenon that is to be explained with the microphysical description or with a combination of analytic statements, is not derivable from the combinations of analytic definitions. Let’s have a look at what would formal structure of the derivation look like in this case.

44. “Qualitative state *R* is the state that plays causal role *C*;

45. Brain state *B* plays causal role *C*;

46. Brain state *B* is qualitative state *R*. (Levine 1999, p. 6)

Here, the premise 44. should be analytic for the derivation to work. However, it is not analytic, although it may be that some qualitative states play certain causal or functional roles, but even if they do it is not the conceptual truth. In Levine’s terms the premise 44. that is supposed to be a binding premise is actually an example of a gappy identity. It is in fact in need of further explanation. It just seems coherently conceivable that there could be a conscious experience that does not play the typical causal or functional role for the related state. No other corroboration for the assumptions is provided.

Now we come to the crucial point of the account. Given that it is coherently conceivable that qualia could not play their typical causal or functional roles, this only shows that the claims about their typical causal or functional roles are not analytic, which suffices only for establishing the explanatory gap. The conceivability argument itself does not demonstrate the metaphysical possibility of causal or functional dissociations. To establish metaphysical possibility in the conceivability argument a new argument would be required.

Now it becomes clearer what it means for the physicalist explanation of phenomenal consciousness to leave out something. Namely, physical-empirical explanations are supposed to use "thin" concepts whose semantics is based on the modes of presentation that bear very little on the cognitive significance in picking out their referents. On the other hand phenomenal concepts almost entirely depend on the "thick" modes of presentation, wherein a property basically presents itself and the mode of presentation is almost exhausted in the cognitive significance. "Thin" concepts leave out the content of "thick" concepts. More importantly, "thin" concepts cannot include the content of "thick" concepts in principle.

Why should this be a problem? Well, that is exactly Levine's point. This situation does not create a metaphysical problem because it is perfectly reasonable that the same property can be picked out in two different ways. For example, we can conceptualize our experiences in terms of the causal roles brain states or properties normally play and, on the other hand, in terms of picking out their immediate qualities. However, claims Levine, this creates an epistemological problem because why the content of a particular "thick" concept is as it is not intelligible, given the content of a respective "thin" concept and also given that both concepts pick out the same thing. Furthermore, exactly because of the relation between the "thick" and the "thin" concepts, the conceivability arguments fail to establish the metaphysical possibility of counterfactual causal, supervenience dissociations or causal or functional inversions. This according to Levine is the price physicalists have to pay if they were to avoid the metaphysical consequences of the conceivability arguments.

2.7 Intelligibility of explanations and causal efficacy

The EG revolves around the notion of gappy identity. There are two main aspects of the problem based on this notion. On the one hand, Levine claims that given that gappiness of the psychophysical identity stems from the incompatibility of thick and thin concepts flanking both sides of the identity statement it is mysterious how that identity could be true. It is so because thin concepts can be formed and applied even without the cognitive significance that is essential for the thick concepts. Thick concepts on the other hand, are

formed and applied regardless of the descriptions of nomic relations of the properties they are picking out and other properties. So even if both kinds of concept had picked out the same property it remains mysterious how that could be true. The other aspect is that the gappiness of the psychophysical identity prevents the inference from conceivability to metaphysical possibility.

This second aspect seems more problematic to me. It concerns the explanations. According to Levine gappiness of psychophysical identity stems from the fact that the bridging premise in the identity statement that is supposed to identify the explanandum with the explanans is in fact in need of further explanation because it cannot be derived from a set of fundamental *a priori* claims as in the case of natural kinds. Having this in mind then, in his view, the conceivability of zombies is another manifestation of the explanatory gap. This latter case seems rather weaker than the former reading. It seems rather elliptical, because we start off without knowing whether the physical and the phenomenal are the same and then we wonder how it could be that they are the same. Whereas in the case of natural kinds we know that the bridging premise is true and it can be derived from a set of *a priori* claims or fundamental claims. That seems odd.

Let's consider the options here. If phenomenal properties were causally inefficacious then there is no way for us to know about them in the first place. This is the standard inconsistency objection against epiphenomenalism. That is fine. Even if we consider them metaphysically distinct from physical properties and still having causal force then we must allow for a situation in which non-physical causes have physical effects. Just by adding the third premise of the causal argument, i.e. physical effects are not always overdetermined, we still end up with the same conclusion, namely physical and phenomenal causes are the same. If they were the same as physical properties and still had physical effects then what is a source of bewilderment? What is left unexplained then?

My point here is that Levine cannot hold both that physicalism is true and that there is an explanatory gap. Regardless of whether we take efficacious phenomenal properties to be fundamental and metaphysically distinct as in Chalmers's sense or the same kind of properties but differently conceptualized, we get the same conclusion when run through the causal argument. Assuming otherwise does not seem to be justified. It seems rather that

these assumptions are intuition driven, and that the underlying intuitions are in fact in need of an explanation.

If this is the case, then, if one wants to maintain physicalism, one must embrace conceptual dualism.

This idea will be thoroughly discussed in chapters three and four. Let's turn now to the most powerful responses to the explanatory gap problem. In the next chapter I will start off with the discussion of the inconsistency objection to the knowledge argument and about what I take to be its major contribution in weakening the explanatory gap argument. After that, I will be examining the so called "phenomenal concept strategy" (Stoljar 2005) in great detail. I will skip other responses to the explanatory gap problem, such as eliminativism, or higher order thought (HOT) theories. Eliminativism will be omitted because I am assuming a qualia realist position. On the other hand, I won't be discussing the HOT and other varieties of this approach simply because the format of this dissertation does not allow me to assemble a complete review of the field.

3. CHAPTER 3

3.1 Phenomenal concept strategy

"The reference-fixing property that is associated with a theoretical concept is identical with the property to which the concept refers. " (Hill and McLaughlin 1999, p. 452)

In our discussion so far I have analyzed and presented several key arguments and notions about the explanatory gap problem. In the first chapter I have discussed the standard anti-physicalist conceivability arguments (Nagel's subjectivity objection, Kripke's modal argument, the zombie argument, the knowledge argument and the two-dimensional argument). In the second chapter I have analyzed Levine's account of the explanatory gap problem (Levine 1983, 1993, 2001). The analysis has teased out two key areas of the problem that require the most of the physicalist's attention. These are the entailment from conceivability to metaphysical possibility and the intelligibility of explanation of the psychophysical identity. In this chapter I would like to discuss the most attractive and the most successful response to all these arguments, an approach that has several varieties but which is commonly known as "the phenomenal concept strategy" (Stoljar 2005).

In its most general articulation the phenomenal concept strategy explains the seeming contingency of psychophysical identity by appealing to a special nature of phenomenal concepts, i.e. by a special way we think about our experiences from the first person perspective. According to the strategy, phenomenal concepts pick out their referents directly, their modes of presentation are substantive and determinate. In this sense, they are unique, irreducible concepts. Furthermore, there are no *a priori* reasons of why phenomenal concepts and physical concepts cannot refer to the same thing. According to this strategy psychophysical identity is necessary and *a posteriori*.

In the following sections I discuss several accounts of the phenomenal concept, which culminate with Papineau's account of the perceptual phenomenal concepts. After

that, in the rest of the chapter I analyze several key issues connected with the phenomenal concept strategy and the explanatory gap problem, i.e. I discuss the "Max Black's objection" to psychophysical identity which claims that distinct modes of presentation of the concepts flanking the identity sign require some distinct properties in virtue of which concepts pick out their referents. The other major issue is the explanatory adequacy of the phenomenal concept strategy, that is, if phenomenal concepts explain the epistemic gap, then they are not physically explicable and vice versa.

I now turn to discussing some of the major accounts of the phenomenal concept strategy.

3.2 Loar's account of the recognitional phenomenal concepts

Unlike physical concepts that pick out their referents via descriptions of causal roles, phenomenal concepts, according to Loar (Loar 2002), pick out their referents directly, as it were, by using perceptual modes of presentation. In Loar's view, natural kinds concepts pick out their referents via contingent modes of presentation. On the other hand, phenomenal concepts refer directly without relying on the contingent modes of presentation. The phenomenal property presents itself in its own right. To get this point across he considers two examples. The first example involves a person, call him Max, who knows that the glass in front of him contains certain chemical compound, say CH_3CH_2OH, without knowing that it is alcohol. Furthermore, he might even lack a concept "alcohol" altogether. What would happen after he learns the concept "alcohol" or after being told that glass of CH_3CH_2OH is actually a glass of alcohol? Loar claims that Max would learn a new fact. He maintains that if we followed the reasoning from the knowledge argument it would turn out that alcohol is not CH_3CH_2OH. It is not that difficult to figure out what is wrong with this conclusion. It is of course the fact that both CH_3CH_2OH and alcohol use contingent modes of presentation to connect to their referent so given that the identity of CH_3CH_2OH and alcohol is *a posteriori* we explain away how it is that Max learns a new fact and why learning a new fact in this case does not imply introducing a new property. In other words, we use contingent modes of presentation as a base for explaining how the same property instantiates two different concepts.

The second example involves Margot, who learns that there is an element in the periodic table called *Au* with the atomic number 79 that people use to decorate themselves. However, Margot has never seen gold and cannot identify it visually. Once she is shown several objects made out of gold and is told that those are samples of *Au* she obviously learns a new fact. If the knowledge argument had a universal validity it would follow that gold is not *Au*, which is obviously false.

These two cases unveil a few very important points about the knowledge argument. The Max case shows that one and the same thing can be picked out by two different concepts without any gap in our understanding of how that could be, if we accept that both concepts use contingent modes of presentation. The Margot case shows that the visual experience in the type demonstrative form and under an opaque context also relies on the contingency that is crucial for the point that the same property can be picked out by two different modes of presentation. In this case the element is picked out on the one hand, via description of the causal/functional roles it standardly plays, and on the other hand, via a type demonstrative that also relies on the contingent mode of presentation: *that type of experience*.

Now consider the knowledge argument. Let's assume that upon release Mary indeed learns something new[19]. It can't be that she learns some old fact under a different contingent mode of presentation like in Max's case, because she does not learn it under a contingent mode of presentation. Put into philosophical jargon, she learns a new fact *simpliciter*. Her knowledge about the new fact is more intimate and direct so there is no contingency in her knowledge about her phenomenal experiences that could be used to explain the fact that she does learn something new. If this is the case, then physicalisim is in trouble because if an identity statement is to be considered as true and based on distinct conceptual semantics, then at least one of the concepts flanking the identity sign has to be

[19] I would like to thank Rob Templing to reminding me of the distinction made in the literature according to which the knowledge argument equivocates on the two different sense of the term "fact": "fact" as "an item of knowledge" and "fact" as "a truth about the world". The point is that physicalism is only committed to making claims about "all facts" in the latter sense, but the knowledge argument only proposes a new fact in the former sense. This is a brilliant point but I shall skip it here because discussing such fine distinctions would require considerably more space than I have in this dissertation and, on the other hand, my analysis here won't be affected significantly by a lack of such discussion.

based on the contingent modes of presentation. This is what Loar calls the semantic premise and of course we know it from the preceding discussions on the epistemological and modal aspects of conceptual semantics, i.e. from the discussions about the relation between apriority/aposteriority and necessity. As we recall, the standard anti-materialist view in this matter rests on the assumption that necessity of *a posteriori* statements stems from the binding topic neutral premise that identifies explanans and explanandum in a meta claim that enables us to evaluate necessity of even *a posteriori* identities based on some underlying a priori claim. I shall be discussing this issue at greater length in the last section of the next chapter within the framework of Block and Stalnaker's (Block and Stalnaker 1999) criticism of conceptual analysis and reductive explanations.

As we have seen in the preceding chapters there are two crucial claims lurking in here. Loar rightly identifies them as the claim of distinct conceptual semantics (or in his terminology: "conceptually independent concepts" (Loar 2002, p. 299)) and a semantic premise. His strategy is to accept the distinct conceptual semantics premise and to deny the semantic premise. Denying the semantics premise comes down to this: if an identity statement consisted of two conceptions, one of which picked out its referents via contingent modes of presentation, and the other conception directly, but both conceptions referred to the same property, then basically the knowledge argument and the conceivability arguments would be ineffective and it turns out that physicalism would be compatible with the distinct conceptual semantics. In other words: if the phenomenal concepts picked out their referents in a direct way, that is, not via contingent modes of presentation, and yet had independent semantics from the natural kinds concepts, so that Mary's history is coherent, then physicalism would be compatible with the knowledge argument. However, there seems to be only one thing in here, not two as Loar seems to distinguish. The fact that phenomenal concepts pick out their referents directly, not via contingent modes of presentation means that their semantics are conceptually independent from the natural kinds concepts. So this is the problem we started with. Let's see how Loar further develops his account.

To do so, let's consider in some detail the semantics of the phenomena concept proposed by Loar. He claims that phenomenal concepts are recognitional concepts. They have the form: *x is one of that kind*. This form also renders them as type demonstratives. They are couched in a disposition to classify objects, events and situations through

perceptual discriminations (Loar 2002, p. 298). We acquire phenomenal concepts directly from our experience but we can also apply them to other situations or we can project/ascribe them to others. For example we say: *she has that kind of experience*. We are able to do this because phenomenal recognitional concepts are based on the dispositions that are connected with the capacities to form images about an identifiable kind even in the absence of the currently perceived instance (Loar 2002, p. 298). Furthermore, recognitinal concepts are embedded in the broader conceptual background of other discriminations. For example, one can utter the following claim: "*a physical object of THAT kind"*, but also the following: "*an internal state of THAT kind*".

But let's go step by step here. In order to fully appreciate this idea let's consider the following: a person goes into a desert and encounters a plant that she hasn't seen ever before. She doesn't know the name of the plant but she is nevertheless able to apply the demonstrative concept of that plant on any future encounters with another plant of that kind. For example, she might say, "Oh, this is that kind of plant". She might be able to do all this without actually having a name for it. This example unveils a very important trait of the recognitional concepts, that is, that the concept can be applied without actually having an instance that was responsible for forming the concept. One has to distinguish phenomenal recognitional concepts from the other forms of recognitional concepts. For example, a person can learn about the chemical composition of porcelain and its causal/functional relations with other compositions or objects without being able to recognize it visually. That is to say, after learning how to recognize porcelain visually, she acquires a recognitional concept under contingent modes of presentation. By contrast the phenomenal recognitional concept is recognitional in its core as Loar puts it (Loar 2002, p. 298), i.e. even when we acquire it, it is recognitional. In another words, they do not need to contain a reference to some past instances of the experience that was responsible for forming a given concept.

Finally, the phenomenal recognitional concepts are perspectival. This means that they are in part individuated by their own constitutive perspective. On the other hand, one could apply these concepts even outside their constitutive perspective. For example, a person sees a four-legged creature from up close and forms a recognitional concept *these creatures 1*. Suppose then she sees some creatures in the distance, but she is not sure whether those creatures are the same kind as the ones she saw from up close, so she calls

them *these creatures 2*. These two concepts are *a priori* independent, one cannot infer one from the other.

Phenomenal recognitional concepts, according to Loar, pick out the internal neuro-functional properties of the brain. We deploy them when thinking phenomenologically about our experiences, but under physical/functional descriptions. So it seems, argues Loar, that there are no *prima facie* reasons nor available philosophical arguments against the claim that phenomenal recognitional concepts pick out those internal neuro-functional properties in the way other than the contingent modes of presentation.

The question immediately arises: is it possible to give some plausible argument that would account for the independence of the phenomenal concepts and the natural kinds concepts? A physicalist would normally say that there is no implausibility involved in the assumption that the truth conditions for both phenomenal concept and a natural kinds concept could be the same and as long as their truth or falsity can be known *a posteriori*. For example, there seems to be nothing *prima facie* wrong with the claims that "*the state A feels like that"* and "*the state A has physical/functional property P"* share the same truth conditions. There is nothing wrong in here because if the phenomenal concepts refer directly, unmediated by the contingent modes of presentation, then the truth conditions that determine any counterfactual situation are satisfied by the physical property picked out by both the phenomenal concept and the physical concept. Someone might object that if both concepts rigidly designate a single property, then it must be the case that they are *a priori* inter-derivable, that is, one should be able to derive the phenomenal concept from the physical concept and vice versa. Loar's response is that this objection conflates two senses of rigid designation here. In one sense rigid designation means something like: "captures the essence of…", which basically means refers directly and unmediated. The other sense reads: "be conceptually interderivable with some theoretical predicate that reveals the internal structure of the designated property" (Loar 2002, p. 300).

However, it is important to note that not all self-directed recognitional concepts are phenomenal concepts. Loar lists two examples for this distinction (Loar 2002, p. 300). The first one involves cramps, involuntary muscle contractions, and the second one involves

cases of blindsight, a phenomenal blindness in restricted retinal areas. In the case of cramps, there is a self-directed recognitional concept, but there is no phenomenal quality to be picked out. The cramps are only muscle contractions, which have characteristic feel, but they are not feelings themselves. As Loar puts it (Loar 2002, p. 300) we can have a recognitional concept of a cramp but it won't be a phenomenal concept for it does not pick out any quality associated with the muscle contraction. Of course we use a phenomenal mode of presentation as he calls it to refer to that feel. In the case of blindsight wherein people are able to "guess" with high accuracy what is in the blind field, although consciously blind in this visual region, we have a self-directed recognitional concept that picks out a "blank" phenomenal concept. Phenomenal concepts on this proposal are a subclass of the self-directed recognitional concepts.

Having characterized the phenomenal concepts in this way, the question immediately arises: if the phenomenal concepts are a kind of self-directed recognitional concepts, then how is it possible that we are able to ascribe them to others?

Loar claims that it is possible if one considers the fact that phenomenal concepts are demonstratives that obtain their reference from one's own case of actually or possibly having an experience of a requisite type. So when we ascribe phenomenal states to the others we use the demonstrative of the following form: *x has an experience of that type* or *x has a state of that sort*. We project to others a recognitional phenomenal concept that actually refers to our own specific physical state. Having said this, another possible objection arises. The other person's physical assembly may differ greatly or slightly from ours, how does the self-directed recognitional concept account for the phenomenal difference given the physical difference? Loar's response is that phenomenal concepts are type demonstratives, so on his view physicalism implies that phenomenal concepts pick out physical ordering, ordering of the physical states, rather than particular physical states. We get phenomenal concepts from our own perspective, remember they are essentially perspectival, and we then apply them to our own states or to the states of others even in the absence of the instance that was responsible for their forming. If we were not able to ascribe phenomenal concepts to others then it would be so because we were not able to apply them in our own case without their original instance or as Loar puts it "they can be ascribed beyond their constitutive perspective" (Loar 2002, p. 303).

Finally, the problem with the explanations and *a priori*/*a posteriori* knowledge that we already characterized in the second chapter rears its head again in here. Namely, when explaining some natural kinds phenomenon in physical or functional terms, we basically analyze it in terms of descriptions of its causal or nomic roles and then show that the physical or functional theory a priori implies that these roles are realized (Loar 2002, p. 304). However, if phenomenal concepts were conceptually independent then it seems impossible to account for an explanation of the phenomenal because no *a priori* analysis is available in this case. Loar's response is that like in the other cases of combining a demonstrative and concepts that are based on the contingent modes of presentation, recall the "CH_3CH_2OH /alcohol" case, do not generate the explanatory gap (Loar 2002, p. 304). Why should it generate a gap in this case? Loar thinks that phenomenal recognitional concepts are not different from the other demonstrative recognitional concepts. He claims that identity statements that consist of the recognitional phenomenal concept and a physical concept are expected to be explanatory, but they are not. He claims that this is a wrong expectation. The intuition that such statements ought to be explanatory is misguiding. This intuition in fact makes it really difficult to see that there is only one property there picked out in two different modes of presentation. However, although the psychophysical identity involving recognitional phenomenal concepts and physical concepts is not explanatory in itself, the sense of puzzlement about how it could be true disappears once we realize the same property can be picked out in two distinct ways. Loar claims that the complete physical description indeed leaves out a subjective conceptions of experience, not because it cannot fully characterize those states, but simply because it does not employ them directly like in the case of contingent modes of presentation.

Now that we have discussed Loar's account of a recognitional concept let's have a look at another version of the same variety of the phenomenal concept semantics, namely Tye's account of the recognitional concept (Tye 2003).

3.3 Tye's account of the phenomenal concept

Tye (Tye 2003, p. 98) suggests that proper phenomenal concepts have to be distinguished from the other sorts of concepts. They are a kind of recognitional concepts exemplified in the following case: I could have a thought that water is wet and recognise that I am having this thought and the act of recognizing the thought is the same as the thought "water is wet", it doesn't add something to the original thought. The very same applies to phenomenal concepts. They are not derivable or inferable from some other concepts. Introspection is a reliable guide for capturing a phenomenal character and applying the phenomenal concept appropriately.

The phenomenal concept strategy is pretty straightforward in answering three major physicalist worries. The conceivability of zombies poses no threat to physicalism if we accept that phenomenal concepts are a different kind of concepts from physical concepts. It is conceivable that there could be a creature physically indistinguishable from me but still lacking phenomenal experiences; however, that is not different from the other kinds of *a posteriori* necessary identities, like "water=XYZ" or that I am not Daniel, etc. The second worry is the knowledge argument. Consider this, if phenomenal concepts were not different from the physical concepts, Mary would have known what is it like to see red colour even before her release. Once we realize that phenomenal concepts are different from the physical concepts it becomes clear that Mary learns something new upon release. However, we need to be extra cautious here not to characterize phenomenal concepts as demonstratives that utilize physical sortals. For example, imagine Mary was even equipped with a cerebroscope, a device that enables one to when that other people are undergoing certain experiences just by pointing the device towards their heads, and when she uses it to look in Mortimer's head she sees and recognizes a physical state that corresponds to Mortimer's seeing red. However, she doesn't acquire a phenomenal concept because she needs to be able to make a "significant" discovery when she herself sees a ripe tomato in order to have a proper phenomenal concept (Tye 2003, pp. 92-3).

Finally, the third worry is about the intelligibility of explanations. To remove this worry there should be no physical concept that is *a priori* co-referential with a phenomenal concept. To appreciate this constraint consider the following example of an explanation:

47. “Pain* is the F

48. Physical state so-and-so is present

49. Physical state so-and-so is the F

50. Therefore Pain* is present.” (Tye 2003, p. 93)

“Pain*” in this argument represents an essence of an unpleasant phenomenal character of pain and “F” is a physical predicate. The second and third premises are clearly empirical ones. The first premise is supposed to be *a priori*. If that is the case, then the conclusion in 50. would follow even if the first premise was not necessary *a priori*. This clearly shows that an explanation using phenomenal concept flanking one side of the identity statement needs no further explanation because the premise 47. is knowable to anyone who possess the concept “pain*”

So for a true physicalist theory of consciousness phenomenal concepts ought to meet three major requirements: a) they have to refer to physical properties but not be physical concepts, b) this, however, does not mean that they are demonstratives that utilize physical sortals, and c) they must not have *a priori* associated co-referential physical concepts. For if they had a priori associated co-referential physical concepts then their reference would be fixed rigidly and we would end up with the same problem again (Tye 2003, p. 93).

To get the point about the imagery feature of phenomenal concepts across consider this example:

51. “the visual experience of red= brain state B” (Tye 2003, p. 100).

The standard anti-materialist view on this example is that when we think about the left hand side of the identity statement we normally conceive its referent as triggering some image of “red” in us. So if the identity is true at all, conceiving of or undergoing a “red” experience triggers the brain state B. On the other hand, when we look at the right hand side of the identity, nothing like that happens. Conceiving of a brain state B does not trigger a “red” image. So on the standard view this seems puzzling.

However, on the account of phenomenal concept just presented, this example only shows that there is a huge difference in the functioning of these two concepts but not in their referent. According to Tye's proposal, phenomenal concepts refer via causal connections they have with their referents (Tye 2003, p. 97). On this account it seems possible for a concept to refer to a phenomenal quality without being a phenomenal concept. For example, imagine a neuroscientist Fred in the future, say in the 22nd century, who has a sensory impediment and is unable to feel pain, say he suffers from the congenital analgesia (Grahek 2007). However, in his time the technology has advanced that much so they have cerebroscopes [20]. When looking through the cerebroscope into someone's brain Fred will be able to form a concept that refers directly to the brain state of a person who has no defects or impediments in their pain experience. Although Fred will be able to feel what that other person feels when in pain and his concept of the other person's experience will refer directly to her experience, he will not be able to acquire a phenomenal concept. He will not possess a proper phenomenal concept because he will not be in a position to know what is it like to have that experience. The upshot of this scenario is that one cannot have a phenomenal concept without knowing what is it like to undergo certain experiences (Tye 2003, p. 98). It also follows from this example that no *a priori* analysis of the concept "phenomenal concept" is possible, i.e. phenomenal concepts are conceptually irreducible. This means that the question about how the phenomenal concept function does not make much sense. For if such a question were meaningful, Mary would not have learned anything new upon release, she would be able to infer the phenomenal facts from the physical facts. On the other hand, the question what makes a certain phenomenal concept a phenomenal concept makes perfect sense. Tye maintains that some characterisation of phenomenal concepts is possible to give. For example, a) they make marks in memory as a result of undergoing certain experiences, b) they can trigger some mental images in response to some cognitive tasks, and c) they enable concept possessors to discriminate between phenomenal qualities that the concept refers to, directly and in introspection (Tye 2003, p. 99). These three features of the phenomenal concepts are what distinguishes phenomenal concepts from other psychological concepts. This is also the reason why Fred the future neuroscientist

[20] See the above explanation of the device

cannot have a phenomenal concept of another person's experiences and this why Mary learns something new upon her release.

What binds these three crucial features of phenomenal concepts together is the assumption that there is an *a priori* link between the concept possession and knowing what is it like and other relevant cognitive abilities, such as recreating an image of experience in memory, etc (Tye 2003, p 99). To make the proposal more picturesque Tye (Tye 2003, p. 101) suggests that phenomenal concepts should be thought of in two different ways. For example, imagine an explorer who discovers a new species of animal and still hasn't got a name for it. He might take a photo of it assign it a name and file it. Next time he wants to talk about the animal he takes out the photo and shows it to his colleagues. The name on the picture refers to a species of animal whereas picture refers to a particular animal, i.e. a token of a type. Another way to make this point is to imagine the same explorer in the same situation, but instead of taking a photo he carves a replica of an animal in the wood. On later occasions he uses the replica both to refer to a type (species) and to a token (particular animal). The same applies to the phenomenal concept, argues Tye. Instead of an animal and a photograph we have a "phenomena character detector" in our heads, that is set up to recognize phenomenal qualities of our experience that occupy the focus of our attention (Tye 2003, p. 102).

Now when the stage is set, by discussing basic ideas behind the phenomenal concept strategy, let's consider its most advanced varieties which will tease out the major shortcomings of the recognitional phenomenal concepts. After that we will be discussing some of the central issues of the strategy.

3.4 Indexical-quotational and perceptual accounts of the phenomenal concept

As we have seen in the previous sections one of the most promising routes to answer the worries raised by the explanatory gap problem is the phenomenal concept strategy. In this section I will be examining the indexical-quotational and perceptual variety of phenomenal concepts and I will show why it is better off than other accounts of phenomenal concepts.

The explanatory gap account posits a problem for physicalism by claiming that the peculiar way we have an epistemic access to phenomenal states determines the way we explain and conceptualize phenomenal experiences. Namely, it makes it seem unintelligible how causal or functional explanation of phenomenal consciousness could be true. The argument provided in support of this claim, as we have seen in the preceding sections, establishes that justification of the identity statements in the case of natural kinds relies on the topic neutral descriptions of the causal roles properties usually play, and ends up with the self-evident truths (Levine 1993, 1998, 2001). However, this structure of an explanation would not work in the case of psychophysical identity because either the topic neutral description of the causal roles is not available in these cases or the justification does not end up with the self-evident truths. It has been argued that the topic neutral description of the causal/functional roles of the phenomenal states is not possible to give because it is difficult to see how phenomenal concepts would refer via a description. This is the point where phenomenal concept strategies step in. Namely, proponents of this strategy concede on the claim that there is a problem if we expect phenomenal concepts to refer via a description and if they do not refer in that way. However, once we recognize the special features of phenomenal concepts, such as direct reference and realize their kinship to the perceptual concepts and demonstrative concepts, then it becomes clear how this strategy addresses the main difficulty of the explanatory gap problem.

Let's elaborate on these claims.

Papineau (Papineau 2002) develops an account of indexical-quotational phenomenal concepts and, later on, an improved account of nondemonstrative phenomenal concepts (Papineau 2007). On his view the phenomenal concepts refer in a direct way by picking out their referents by focusing on the immediate quality or "the feel" of experience, in another words by indexing the requisite experience, or by "quoting" it by recreating it in imagination. In this way the structure of reference of phenomenal concept takes up the following form: *the experience:*__. The underlined space is then filled either with currently undergoing experience or by some past experience recreated in imagination. Understanding phenomenal concepts in this way then justifies the claim that phenomenal and physical concepts although referring in two fundamentally different ways still pick out the same property, one in terms of its causal roles the other one directly in terms of its immediate

quality. Papineau calls this position conceptual dualism; that is, there is only one kind of property ontologically but two different ways of how we grasp them conceptually. In other words, ontological monism and conceptual dualism are two compatible doctrines about the world.

On Papineau's view, this variety of the phenomenal concept does not only answer the worries raised by the knowledge argument and by the conceivability argument, but it also explains why dualist intuitions seem so compelling. Furthermore, this account of the phenomenal concept does indeed answer the explanatory gap argument in a way that Levine exactly requires (Levine 1983, 1993, 2001). Namely, Papineau's account of phenomenal concept meets two main requirements for closing the gap: 1) it does explain away why it seems that the psychophysical identity leaves us with the explanatory gap and 2) it shows that justification of the psychophysical identity does not have to rely on the *a priori* analysis of meaning [21]. Let me expend a bit here on both points.

As we have seen from the example of the structure of explanation in the case of natural kinds identities, the justification of the identity claim does not show why for example water is H_2O, because if they are identical then it doesn't make sense to ask why they are identical. In another words, if someone understands what identity is, then the question why A is identical to B doesn't make sense. This is what being identical means. It is especially clear in the case of proper names in Kripke's theory of meaning. If Cicero is Tully, then it doesn't make sense to ask why Cicero is Tully. There are only two names that denote one and the same person, because according to Kripke, proper names refer directly. The same point can be applied to physical and phenomenal concepts, argues Papineau. We have one property that is referred to in two different ways. One is by picking it out directly (phenomenal concept) the other one by describing causal or functional roles given property may play (physical concept). On the other hand, the justification of the natural kinds identities tells us why, for example, water has some further properties like being odourless, tasteless, colourless, etc. Once it is established that a statement is an identity statement, it ceases to make sense to ask why its constituents are identical. We will come to this issue in

[21] I shall be discussing this issue much more thoroughly in the next chapter

the next chapters when we will be discussing in greater detail the difference between natural kinds and psychophysical explanations, especially Block and Stalnaker's (Block and Stalnaker 1999) analysis of the reductive explanations.

Let's return now to the issues from the beginning of this section. It has been said that conceptual dualism (Papineau's account of phenomenal concepts) and property dualism claim that phenomenal concepts refer in a direct way and not by picking out causal roles of the properties. The conceptual dualism claims there is only one property under two distinct modes of presentation, whereas property dualism claims there are in fact two properties. Papineau provides independent argument that corroborates this idea. It is the causal argument for physicalism. As we recall from the introduction, the argument goes like this:

1. Conscious mental occurrences have physical effects.

2. All physical effects are fully caused by purely *physical* histories.

3. The physical effects of conscious causes aren't always overdetermined by distinct causes.

4. Therefore, materialism is true. (Papineau 2002, pp. 17-18)

This argument brilliantly and elegantly encapsulates the very idea of physicalism. Premise 3. in this argument guarantees (by preventing overdetermination) that the effects of conscious and physical causes are not distinct causes. Premise 1. claims that certain effects have conscious causes. The premise 2. claims that certain effects have physical causes. If conscious and physical causes are not distinct, as the premise 3. claims, then the only reaming possibility is to conclude that conscious and physical causes are identical.

Consider one more time the knowledge argument. If Mary learns something new upon her release then the knowledge she acquires is about some properties that are distinct from physical properties, for if they were not distinct she would have known them already before the release, wherein she knew all the physical facts. This is Jackson's reasoning about the knowledge argument, which Chalmers also adopts in his work. However, there is nothing in the knowledge argument that justifies this intuition. On the contrary, it rather seems as if the intuition of distinctness of properties stems from what Papineau calls the antipathetic fallacy. Papineau argues that the two ways in which we acquire knowledge about

experiences, namely the first person and the third person, fuel the intuition that there must be to distinct properties. It seems as if the first person perspective involves the experience itself whereas the third person perspective leaves out this very experience. He calls it that way because it seems that in arguing against materialism people just fail to ascribe two distinct modes of presentation to the one and the same property and instead claim that there are two kinds of properties. In coining this name he draws on what is known in poetry as the "pathetic fallacy", that is, a fallacy of ascribing human feelings, beliefs or attitudes to the natural phenomena. For example, we find the "pathetic fallacy" in these sorts of statements: "gloomy forest", "punishing sun", etc.

However, isn't it equally plausible to assume that Mary does in fact learn something new, but it doesn't add anything to the ontology? Doesn't it seem even more plausible that she rather acquires a new concept, which refers to the same property as her material concepts but under a different mode of presentation? It seems as if she does. Even Jackson changed his mind about the consequences of the knowledge argument. He now claims that the knowledge argument does not show that what Mary learns upon release implies that there is a distinct property, but rather that what she acquires can be explained in terms of representationalism about sensory experience. I will elaborate on this point in the following sections on the inconsistency objection.

Let's return now to the semantics of phenomenal concepts.

In order to develop the structure of reference on his account of phenomenal concepts, Papineau analyzes two kin sorts of concepts: perceptual and phenomenal concepts. He argues that the perceptual concepts are those of the sensory experiences. They are like sensory templates that perhaps even animals have, and are ready to be used in situations that trigger the requisite experience. Loar would probably call them recognitional concepts. The latter sort seems to be unique only to the linguistic creatures. Let's elaborate on all of these distinctions.

Although perceptual concepts in their linguistic form do contain the demonstrative "that", and thus can refer to different things at different times, it would be wrong to categorize them as demonstratives because one might use the same linguistic expression (demonstrative formula) to refer to the same thing every time one encounters the same

thing. Papineau gives an example of a bird at the bottom of a garden. First time he sees it he might say, ”Oh, I like the song of *that* bird”. Next time he sees it he still doesn’t know what kind of a bird it is or he doesn’t have a name for it, he would then use the perceptual concept stored in his memory from the previous encounter with the bird to refer to the same bird, but still using the demonstrative linguistic form. He would say, “Oh, look here is *that* bird again”. He perhaps might not be able to categorize or name it explicitly, but he will be able to refer to it by using a perceptual concept which is formed by attaching distinctive information about the perceptible item (a bird in this case) to the sensory template. The upshot of this idea is that although perceptual concepts may take up the linguistic form of a demonstrative they should not be considered as demonstratives because the perceptual concept itself is like a stored template of the sensory experience that can be reactivated every time an agent undergoes the requisite experience. In a sense, the stored template will be “activated” by the experience it refers to. However, the question about how we know whether these sensory templates refer to a type or a token arises straight off. Papineau’s answer fits plausibly to the whole theory. He says that if it is a particular bird that triggers some template then the template refers to a token, if ANY bird of the species activates some template then it refers to a type. This question emphasizes some inevitable difficulties with the recognitional account of perceptual concepts, namely, that it does not provide sufficiently fine-grained referential distinctions that are required for different contexts. Purely recognitional account of perceptual concept does not allow for misrecognition, that is it does not allow for a sensory template to be misapplied. It is compatible with the cases in which, for example, two persons might have the same sensory template but one person thinking about a type and the other person thinking about a token. On the other hand, Papineau’s version of the perceptual concept shifts the focus from the use to the function of the concept. The gist of this idea is that what determines the semantics of a concept is not how it is applied across different contexts, but rather what kind of information is attached to the sensory template. So it seems that the origin of a perceptual concept and the kind of information that gets attached to it determines how it refers in different contexts. One can think of attaching the information to the sensory template as filling the “slots”. For example, on an encounter with a bird with a missing claw a subject stores that information in some “slot” of the sensory template. Next time the subject encounters that bird she will apply the perceptual concept to that particular bird, not to the whole species. On the other hand,

when she sees any bird eating seeds, she would store that information in the template and apply it to the whole species. The main point to be noted here is that a perceptual concept normally refers to its origin, but not to whatever we believe it fits.

Papineau says that the reason why we use nondemonstrative perceptual concepts in the demonstrative linguistic form is because sometimes we simply don't have a name for the thing we are referring to. However, on this account when the thing that we want to refer to by using nondemonstrative perceptual concept is not present, we find ourselves unable to convey the thought. This is what is meant by the claim that perceptual concepts are stored sensory templates. They are set up on the first encounter with their referents and they can be reactivated on every other encounter. However, they can also be reactivated even in the absence of the relevant experience. This is the case that exemplifies earlier mentioned recreation of a perceptual concept in imagination. We might reasonably assume that we share this concepts of this kind with other animals, and that their purpose is to guide creatures in everyday survival. By the same token, we can assume that various bits of information about the referent will be attached to these templates and will be activated when the template is activated in both encountering the referent or imaginatively recreating it. This function of carrying information and adding new features to the template is critical in distinguishing perceptual concepts from demonstratives, because demonstratives don't carry or add such information. They only refer to the item in a given context independently from the previous encounters. However, these repositories of information that are attached to the sensory template actually enable perceptual concepts to refer to the same things in different contexts. In a word, this amounts to showing that the function of a perceptual concept is to accumulate information about the item that was responsible for its formation. Now, given that the function of a perceptual concept is understood in this way, it becomes clear how this account allows for a misapplication of a concept.

Understood in this way, perceptual concepts shift the focus of reference from the phenomenal nature to the cognitive function. That is, instead of capturing whatever it is like to have an experience, their semantics is now turned towards how we form concepts of our own experiences that guide us in everyday situations. In this way a phenomenal concept is partially constituted by the "what-is-it-likeness" of experience. Furthermore, the phenomenal concept strategy understood in this way adds another important dimension to

the approach, namely the empirical commitment. Of course, according to this strategy the phenomenal concept is experience dependant, that is, their acquisition depends on undergoing the requisite kind of experience. Mary needs to see the red colour on her own in order to have a concept of what is it like to see the red colour. This is where the inconsistency objection steps in. I will be discussing this idea in the next section. Let's stick a bit longer with the semantics of the phenomenal concept.

The same idea about activating attached bits of information in the slots of the sensory template either by actually perceiving an item that formed that template or recreating it in imagination, applies to the concepts that allow us to think or to refer to a perceptible item even when it is not actually perceived or recreated in imagination. Papineau calls these concepts: *perceptually derived concepts*. When a sensory template is formed, along with the slots for information about the referent, some other "file" is opened that shares some of the slots but that is not activated by a perceptible item or by a imaginatively recreated perceptible item. To put it differently, these perceptually derived concepts allow us to think about perceptible items even when they are not actually present, both in experience and in imagination.

Finally we come to Papineau's improved account of phenomenal concepts.

Phenomenal concept, according to Papineau, is a special case of a perceptual concept. One can think of the concept as a sensory template deployed to think about experiences themselves, not about objects of experience. We explained earlier how it is possible that perceptual concepts can refer to types and tokens. It was said that when the slot of a sensory template is filled with the species specific information then the perceptual concept refers to a type. When it is filled with the individual specific information then it refers to a token. The same line of thought can be applied to phenomenal concepts. When a perceptual concept is deploying experience specific information then we have a phenomenal concept. However, the referential distinction between types and tokens of experience cannot be made in the case of phenomenal concepts. It rather seems as if phenomenal concepts always refer only to types of experience. Whenever we try to refer to particular experiences we find ourselves deploying some more sophisticated concepts that are referring perhaps via a description. The upshot of this idea is that we can think

phenomenally about experiences and then we are deploying some basic perceptual concept in combination with some more sophisticated concept that refer via descriptions to particular experiences. On the other hand, we can think phenomenally with experiences by deploying a perceptual concept.

In the following sections I will be examining several key issues of the phenomenal concept strategy and thus setting up the stage for the next chapter wherein I will be presenting my key argument to the effect that the phenomenal concept strategy is the right approach to solving the explanatory gap problem. In addition, I will be considering some epistemological issues of explanations of natural kind identities and psychophysical identities in the framework of criticisms of *a priori* physicalism and in the light of the debate between Block/Stalnaker and Chalmers/Jackson.

Let's see now whether there is some independent argument for conceptual dualism apart from the corroboration it gets from the causal argument for physicalism.

3.5 The inconsistency objection and the phenomenal concept strategy

As we recall from the first chapter the knowledge argument revolves around the idea that either Mary learns something new upon her release or not, and if she does learn something new then the knowledge she gains is of a different kind then the knowledge about the physical basis of colour vision. For if that knowledge were not of a different kind, she would have known or have been able to infer it from the body of physical information. This is then taken as a starting point in an argument that claims that the fact that she learns something is strong evidence that phenomenal properties exist, but since she did not know those facts or was not able to infer them from the physical facts while still in the room, indicates that they are causally inert, so epiphenomenalism must be true.

I think the knowledge argument does not show any of that without introducing tacitly some additional assumptions. I particularly have in mind the assumptions about the modes of presentation of natural kinds concepts and concepts of the experience from the first-person perspective. It seems that Jackson and others believe(d) both *a*) that concepts picking out physical properties use only modes of presentation based on the descriptions of

the causal roles properties normally play, and *b*) that the difference in modes of presentation is a manifestation of the metaphysical difference in properties.

In this section I would like to argue against both of these assumptions and furthermore to prove that if she were not able to learn anything new it would pose a much more severe problem for both materialists and anti-materialists. In the years after the Knowledge Argument had been published several other arguments against it also appeared. I refer in particular to the arguments by Campbell (Campbell 2003), Watkins (Watkins 1989), Stjernberg (Stjernberg 1999) that focus mainly on the claim that epiphenomenalism and the claim that Mary learns something new are mutually exclusive. The general structure of these argument goes roughly like this: if she were able to learn something new then it means that phenomenal properties were causally efficacious, and thus epiphenomenalism cannot be true. On the other hand, if epiphenomenalism were true she should not be able to learn something new, as phenomenal properties then would not have any effects, and therefore we could not know about them. This obviously isn't enough to block the Knowledge Argument completely, but it makes a very good point upon which I want to build my argument in this section. In particular, the point I want to argue for in this section is that this kind of analysis of the Knowledge Argument makes a good case for the phenomenal concept strategy.

In most of these arguments it is argued that the Knowledge Argument fails to show that epiphenomenalism is true although it is conceded that the argument establishes that qualia are non-physical (Campbell 2003). In other words, if epiphenomenalism were true then Mary should not have learned anything new upon her release, because if phenomenal properties were causally inefficacious then we would not be able to detect them, neither from the first person perspective nor from the third person perspective. However, in this case the very idea of epiphenomenalism would not make any sense because it does not prove that phenomenal properties exist in the first place. The only remaining backup for the claim that she learns something new stems from our intuition that she "must" learn something new; it seems so obvious that she learns something new. On the other hand, what is the use of a theory that is not able to detect the relevant properties by failing to bring them into nomic relations with other properties from the same plain or physical properties?

It is also claimed by the proponents of the inconsistency objection that although the Knowledge Argument fails to prove the truth of epiphenomenalism while holding phenomenal properties causally inert, it succeeds in showing that they are not physical. The argument itself does not show that they are not physical without an additional tacit assumption that phenomenal concepts refer only via descriptions of the causal roles. The whole fourth chapter will be devoted to this issue, so I won't be discussing it at length here.

I think that the materialists must be committed to the view that Mary learns something new upon her release because if she hadn't learned anything upon release it could only mean that we would not be able to know about phenomenal properties, not even from the first person perspective.

If it is taken to be logically possible that phenomenal properties are causally or functionally inert, as proposed in the epiphenomenalist argument, then a nomological connection between them and physical properties cannot be established, and therefore empirical evidence for psychophysical laws and the first person phenomenal knowledge itself are jeopardized. This leaves only one alternative: the Knowledge Argument leads to inconsistency in assuming causal inertness of phenomenal properties.

This reason is grounded on the analysis of the anti-functionalisation argument given by Michael Pauen (Pauen 2006), which tries to show that it would not be possible, in principle, to verify a nomological connection between phenomenal and physical properties, if we accept the main premise of the Knowledge Argument. His argument is based on a disconfirmation constraint, a principle that tells us that the experiment should be able to provide counterevidence in a counterfactual situation where the starting hypothesis is false/wrong in principle. The assumption of possible causal inertness of phenomenal properties violates this principle; hence, empirical evidence of their existence cannot be confirmed. Furthermore, there is no way in which psychophysical dissociations, like inverted qualia or absent qualia, could be confirmed nor ruled out, if phenomenal properties were causally inert. Likewise, "our first-person knowledge of phenomenal states is jeopardized if we assume that mental properties are causally inert" (Pauen 2006, p. 16).

On the other hand, and this is the gist of my analysis here, the very fact that Mary learns something new only upon her release and that properties that she learns about could

not be causally inert supports the claim that the concepts she acquires are based directly on experience and also that they do not refer via modes of presentation that are based on the descriptions of the causal roles properties normally play. If it were otherwise we would not even be able to detect those properties or they would have been contained or derivable from the body of physical information.

3.6 Some issues with the phenomenal concept strategy

3.6.1 The Max Black's infamous objection and modes of presentation

Ned Block claims (Block 2007) that at its core, the problem of the explanatory gap is about the modes of presentation. He claims that if we distinguish between cognitive modes of presentation and metaphysical modes of presentation and recognize their dynamics in various versions of the argument for property dualism, we will see that those arguments pose no threat to the physicalism and that conceptual dualism can be true. Let me first make the main move assumed in Block's argument and then I will fill in the details on the way. Block starts off with the claim that is known as the Max Black's objection (Smart, 1959). Basically Max Black objected that if one is to postulate an identity between the phenomenal and the physical then the concepts flanking the identity sign that deploy independent modes of presentation are based on distinct properties in virtue of which these modes of presentation differ. What seems to be the problem here is that phenomenal concept needs to have a phenomenal mode of presentation which ought not be physical, that is, it should not be the property itself. Block argues that this view stems from the standard Fregean approach to reference. For example, consider the identity statement "Hesperus=Phosphorus". Each concept flanking the identity sign has its own physical mode of presentation, say Hesperus's mode of presentation is a property of appearing in the sky in the morning at certain time. On the other hand, Phosphorus's mode of presentation is some other physical property, say a property of appearing in the sky in the evening at certain time. This is what constitutes the metaphysical mode of presentation according to Block. However, there is also the cognitive mode of presentation involved here, which is to be understood in the standard Fregean way as cognitive significance. The cognitive significance is used to explain why some identity statements are informative and why some other are not. For example, the statement "Archie Leach is Carry Grant" seems more informative than

“Archie Leach is Archie Leach”. According to the Fregean view the former statement is more informative because of the tacit knowledge about the formal relations underpinning the identity relation.

According to the standard anti-materialist view, psychophysical identity ought to have metaphysical and cognitive modes of presentation as in standard identity statements in order to be true. Contrary to this view, Block (Block 2007) argues that it shouldn't be the case with the psychophysical identity.

The anti-materialists seem to assume that the difference in cognitive modes of presentation entail the difference in the metaphysical modes of presentation. However, this does not have to be the case. Block (Block 2007) uses a brilliant example to show that the difference in cognitive modes of presentation does not entail the difference in metaphysical modes of presentation even in cases where it seems straightforwardly so. He asks us to imagine a student of French language at the beginner’s level. One day his teacher shows him a cat and explains that a French word for “cat” is “chat”. However, his absentminded teacher later on the same day shows him the same cat and introduces the same word “chat” again. The student then tacitly assumes that there is some biological or physical difference between the French word “chat” associated with the first showing of a cat and the French word “chat” associated with the second showing of the cat. But there is no relevant biological or physical difference between two occurrences of “chat”/cat. We might say that there is no difference in the metaphysical modes of presentation between two occurrences of “chat”/cat. There is only a difference in cognitive modes of presentation. However, our student might think that there is a difference in the metaphysical modes of presentation, whereas in fact there is no such difference in this case. This example strongly suggests that the difference in the cognitive modes of presentation does not entail the difference in the metaphysical modes of presentation.

The main strategy for Block (and for Loar as well) is if there is no plausibility to the claim that cognitive modes of presentation implicate metaphysical modes of presentation, then physicalism skips the modal argument pitfalls. That is, the difference in cognitive modes of presentation does not entail the difference in metaphysical modes of presentation, thus conceptual dualism is not jeopardized by the conceivability arguments.

Let's discuss now another very important issue connected with the phenomenal concept strategy. This is the explanatory adequacy of phenomenal concepts.

3.6.2 Explanatory adequacy of the phenomenal concept strategy

As we have seen in the previous section, some of the main problems for the phenomenal concept strategy are Max Black's objection (Block 2007) and the intelligibility of the explanation of the phenomena concept strategy (Chalmers 2007).

We have already discussed Max Black's objection, so let's turn now to the issue of the explanatory adequacy of the phenomena concept strategy.

Chalmers claims that the phenomenal concept strategy has to be committed to three basic claims:

52. Human beings have a certain key feature *C* such as phenomenal concepts;

52. This feature *C* explains our epistemic situation in regard to different epistemic gaps, that is phenomenal concepts explain the explanatory gap;

54. Finally *C* is explicable in purely materialistic terms. (Chalmers 2007, p 172)

He argues that physicalists cannot simultaneously hold thesis 53. and thesis 54. as true. That is, if phenomenal concepts can explain why there is an epistemic gap, then they cannot be explained in physicalistic terms and vice versa.

For Chalmers this argument is based on the link between conceivability and explanation, instead on the link between conceivability and metaphysical possibility. As we will shortly see, this maneuver does not help much his argument because it is basically an old argument under a new guise, which is already answered by the phenomenal concept strategy. His new argument then goes like this (where "P" represents physical duplicate and "C" represents phenomena concepts) :

55. (P&~C) is conceivable, or (P&~C) is not conceivable;

56. If (P&~C) is conceivable, then C is not physically explicable.

57. If (P&~C) is not conceivable, then C cannot explain our epistemic situation.

---

58. Either C is not physically explicable, or C cannot explain our epistemic situation. (Chalmers 2007, p 174)

Basically, Chalmers's master argument against the phenomenal concept strategy is very similar to Levine's account of the explanatory gap problem (Levine 2001). Whereas Levine (Levine 2001) claims that conceivability arguments fail to justify an inference from conceivability to metaphysical possibility of zombies, they do so at the price of a gappy identity, thus opening the explanatory gap[22]. Chalmers's on the other hand argues that either the phenomenal concept are physically explicable, but that explanation does not help us understand why there is the epistemic gap, or the phenomenal concept strategy explains why there is the gap but it is not itself physically explicable. The trouble for the proponents of the strategy is that for the approach to be true, both physical explanation of phenomenal concepts and explanation of our epistemic situation based on the phenomenal concepts, have to be true in his opinion.

Again, as we have seen in the first chapter, Chalmers believes that the only explanation of why there is the epistemic gap, that is how is it possible for two concepts that use different modes of presentation to refer to the same property, is because the difference in modes of presentation is based on the metaphysical difference in properties. Clearly, if the only way to explain the epistemic gap is through the metaphysical difference in properties phenomenal concepts pick out, then of course they cannot have a physical explanation.

Let's go now into the details of Chalmers's master argument. As always, his argument takes the form of a conceivability argument. He asks whether it is conceivable that there are

---

[22] For the details see the chapter 2

creatures, say zombies, that lack the key feature of consciousness such as phenomenal concepts. At this point his argument is not yet based on the metaphysical possibility of zombies. There are only two possible answers to this question: either it is conceivable that they lack phenomenal concepts or it is not conceivable. This allows for setting the argument in a form of a dilemma.

Let's see how he defends his two premises.

The support for the first premise stems from the considerations about the link between conceivability and explanation, not as one might have expected from the link between conceivability and metaphysical possibility. It seems to him that if one can conceive of (P&~C) it means that there cannot be an explanation of phenomenal concepts in wholly physical terms that makes it transparent why phenomenal concept strategy is true. More precisely, the relationship between conceivability and explanation here is based on Chalmers's account of reductive explanation: to explain something reductively means to make it transparent why certain higher-level truths obtain given the lower-level truths. If it were conceivable that some lower-level truths obtain without higher-level truths obtaining then the reductive explanation fails. Thus, the original explanatory gap persists (Chalmers 2007, p. 174).

The second horn of the dilemma says that if (P&~C) is not conceivable then phenomenal concepts do not explain the explanatory gap. The argument for this claim relies on the same kind of reasoning as the argument for the first premise. It goes like this:

"59. If (P&~C) is not conceivable then zombies satisfy C;

60. Zombies do not share our epistemic situation;

61. If zombies satisfy C but do not share our epistemic situation, then C cannot explain our epistemic situation;

---

62. If (P&~C) is not conceivable then C cannot explain our epistemic situation". (Chalmers 2007, p. 176)

Chalmers holds that premise 60. is the drive of the argument. He claims that the premise 60. amounts to the following claim: (P&~E) is conceivable, where *"E"* stands for our epistemic situation.

However, premise 59. seem very much like a standard claim about the entailment from conceivability to possibility. It says that if it is not conceivable for the physical facts to hold without phenomenal facts holding then zombies have phenomenal concepts. In another words, if an explanation of phenomenal concepts is not *a priori* derivable from the physical facts, then it is metaphysically possible for zombies to have phenomenal concepts or some corresponding beliefs. Understood in this way, his argument brings us back to where we started and the phenomenal concept strategy applies again.

Let's follow Chalmers in setting the argument that summarizes his overall approach here:

"63. (P&~E) is conceivable;

64. If (P&~E) is conceivable, then (*P&~C)* is conceivable or (*C&~E)* is conceivable;

65. If (*P&~C)* is conceivable, *P* cannot explain *C.*

66. If (*C&~E)* is conceivable, *C* cannot explain *E*.

_____

67. *P* cannot explain *C* or *C* cannot explain *E*" (Chalmers 2007, p. 179). The conclusion follows: either phenomenal concepts cannot be explained physically or phenomenal concepts do not explain our epistemic situation.

Here we are again at the same point that iterates throughout anti-physicalist arguments. The reason why it seems that there seems to be a disrupted link between conceivability and explanation is because anti-physicalists tacitly assume that either for phenomenal concepts to be explained physically Max Black's objection will always apply – i.e. cognitive modes of presentation will entail metaphysical modes of presentation, or if the phenomenal concepts are explainable physically they need to rely on the modes of presentation that are based on the descriptions of the causal roles, thus failing to explain the explanatory gap. However, as we have seen in our discussion on the metaphysical and cognitive modes of presentation, this does not hold. To begin with, this disjunction does not follow from Chalmers's argument alone, instead it is tacitly assumed as one of the premises, and as far as I can see there is no independent argument for this claim anywhere in his paper. There is an attempt to justify it through considerations about a link between a priori knowledge and explanation in the famous debate between Chalmers & Jackson and Block & Stalnaker, but as we shall see later on it fails.

Diaz-Leon (forthcoming) makes a very similar point in criticizing this argument. She claims that the task of the phenomenal concept strategy is not to explain our whole epistemic situation, but only why phenomenal concepts are not *a priori* entailed from a complete physical story. According to this interpretation then, the question should be whether zombies that possess relevant phenomenal concepts could infer phenomenal truths from the physical truths? Actually, there are two questions here: one is whether zombies that possess phenomenal concepts could infer a priori phenomenal truths from the physical truths, and the other is whether zombies possessing their own corresponding "zombie phenomenal concepts" could infer *a priori* phenomenal truths from the physical truths. The first one does not make much sense because zombies by definition cannot have phenomenal concepts, so that case cannot be veridical. The second question comes down to the original question of whether phenomenal truths are *a priori* inferable from the physical truths. Of course, as I have pointed out earlier, phenomenal concepts refer directly, not via contingent properties so the psychophysical identity remains necessary and yet *a posteriori*. What Chalmers seems to imply, according to Diaz-Leon, is that the phenomenal concept strategy has to explain conceivability of zombies in terms of them not having states of which we have

a substantive and determinate knowledge. In another words, there should not be any epistemic gap in order for the phenomenal concept strategy to explain the explanatory gap. This seems rather excessive and problematic. Diaz-Leon maintains that Chalmers mischaracterizes the explanatory gap and the phenomenal concept strategy. This is very much in line with my view of the explanatory gap and the overall approach of the phenomenal concept strategy.

Another venue of responding to Chalmers's argument we find in Balog's forthcoming paper (Balog forthcoming). Balog (forthcoming) argues that Chalmers's argument can be laid out in much more detail that would more precisely reflect physicalist commitment with the phenomenal concept strategy. She unpacks the relevant premises according to the distinction between phenomenal and physical characterizations of the phenomenal concepts. In her view, the argument should go like this:

"1Phen) If P&~CPhen is conceivable, then CPhen is not physically explicable.

1Phys) If P&~CPhys is conceivable, then CPhys is not physically explicable.

2Phen) If P&~CPhen is not conceivable, then CPhen cannot explain our epistemic situation.

2Phys) If P&~CPhys is not conceivable, then CPhys cannot explain our epistemic situation." (Balog forthcoming, p. 12)

One set of premises in here is vacuous. This is the set of premises 1Phys) and 2Phen). Premise 1Phys) basically says that if it were conceivable for the physical facts to hold without phenomenal concepts (under physical characterization) holding, then phenomenal concepts are not physically explicable. This is trivially true because if phenomenal concepts were physically explicable then it would follow *a priori* that P&~CPhys is not conceivable. The premise 2Phen) is also already implied by the phenomenal concept strategy. It basically says that zombies without phenomenal concepts are conceivable. This is one of the main tenets of the phenomenal concept strategy. Both claims are envisaged and contained in the

phenomenal concept strategy. So the question arises now whether the other set of premises, namely 1Phen) and 2Phys) are capable of undermining phenomenal concept strategy?

Balog (Balog forthcoming) argues that the other combination of premises, namely 1Phen) and 2Phys), is nontrivially true. Furthermore, she holds that the only way to successfully explain our epistemic situation towards consciousness is by the phenomenal concepts under phenomenal characterization, i.e. C-phen. However, this characterization of phenomenal concepts is not physically explicable, which on the other hand she is willing to admit. That is, this combination of statements does not pose any threat to physicalism. Let's see her argument for this claim.

"1Gap) If there is an explanatory gap between *P* and *C* then *C* is not physically explicable.
2Gap) If there is no explanatory gap between *P* and *C* then *C* cannot explain our epistemic situation " (Balog forthcoming, p. 17).

This is not a further criticism of the phenomenal concept strategy; this is merely restating the original explanatory gap argument already given in (Levine 2001). 1Gap) is the point which all agree with. It seems reasonable to assume that if C were physically explicable, there would be no explanatory gap. Again, phenomenal concepts would have been physically explicable and there would be no explanatory gap if phenomenal concepts used modes of presentation based on the descriptions of causal roles their referents normally play. Chalmers does not provide any additional or independent argument for these two claims so I must conclude that his argument fails to show that phenomenal concepts do not account for the explanatory gap.

His argument is very important for another reason. It accentuates the importance of the issue of *a priori* knowledge and explanation in special cases such as psychophysical identity. I will be dealing with it in the next chapter where I will also provide my key argument for the type B materialism.

In this chapter I have discussed the phenomenal concept strategy as the most

plausible answer to the conceivability arguments and to the explanatory gap problem. I have analyzed the most relevant accounts of the phenomenal concept's semantics, starting from the recognitional concept through indexical quotational and finally, as the most developed account, Papineau's phenomenal concepts as a subclass of perceptual concepts. Apart from that I used the inconsistency objection to the Knowledge Argument (the claim according to which epiphenomenalism and the causal inertness of phenomenal properties cannot be held together) to argue that the phenomenal concept strategy is the best answer to the conceivability arguments. In the last two sections I have discussed two major issues of the strategy, i.e. the "Max Black's objection" and the explanatory adequacy of the phenomenal concepts and concluded that neither of these two objections manages to undermine the phenomenal concept strategy.

In the next chapter I will present my key argument against the explanatory gap problem, and in the light of it discuss the issue of *a priori* knowledge and intelligibility of explanation of the psychophysical identity.

4. CHAPTER 4

4.1 The intuition of distinctness: how to resolve a brute disagreement

"(Perhaps the most significant contribution of experimental philosophy, then, is to underscore the likelihood that)… various intuitions are not theory neutral, but products of such theoretical commitments. If intuitions aren't widely shared by the folk and fit conveniently with theories espoused by those who invoke the intuitions, it's natural to ask whether the intuitions encapsulate those theories, rather than providing independent support" (Rosenthal 2010, p. 5).

After three chapters of setting a stage I am now in a position to lay out my own account against the conceivability arguments and the explanatory gap problem. Let me first recapitulate what has been discussed so far to make my line of argument clearer.

One of the central and vexing philosophical issues concerning consciousness is whether there could be an intelligible explanation of the relation between what it is like to undergo certain experiences and neural processes involved in those experiences (Nagel 1974; Levine 1983, 1993, 2001). This issue is commonly referred to as the *explanatory gap problem.* On the other hand, there are philosophers who claim that such an explanation cannot be made intelligible in physical terms, and furthermore it is a manifestation of a metaphysical difference in properties (Jackson 1982, 1986; Chalmers 1996). This position is commonly known as the *ontological gap*. The best response to both gaps in my opinion is the *phenomenal concept strategy*. The proponents of the strategy concede the conceivability of zombies and that Mary learns something new upon her release, but deny that zombies are metaphysically possible and that what Mary learns is based on an additional non-physical property in the world. These claims are usually based on a particular account of phenomenal concepts according to which they refer directly and not via descriptions of the

causal roles properties normally play. But the most important reason for maintaining conceptual dualism is the causal argument. They argue that psycho-physical identity is a proper identity that does not require further explanation, as with all proper identities, and they further claim that the a posterioriti of psychophysical identity can be explained away by means other than the modal semantics. For example, Papineau (Papineau 2002) uses the notion of the *antipathetic fallacy* to explain the intuition of distinctness that in fact drives the anti-physicalist arguments[23]. I am pretty much on the same page with Papineau as regards this issue.

However, I would like to sharpen a distinction that I believe is not sufficiently appreciated in the literature and that can help me better express what I think is the problem here. It is the distinction between an epistemic gap and an explanatory gap. What is the difference between an epistemic and an explanatory gap? We have an epistemic gap when there are two or more ways of knowing something that are independent from each other, i.e. one can know one thing in one way but cannot know the other thing in the same way and vice versa. The explanatory gap arises when we cannot explain the epistemic gap. Here is an example of the epistemic gap we find in the domain of consciousness. We can know certain things objectively, that is, regardless of the perspective. For that we only need to know the descriptions of the causal roles normally played in this domain. For instance, we can know in this way that a person undergoes certain visual experiences. What we cannot know in this way is what it is like for that person to undergo those experiences. On the other hand, knowing what is it like to undergo some experience from our own first person perspective will never enable us to know the descriptions of the casual roles that properties of our own experience play. More intuitively: we can introspect as much as we want, but that alone will never bring us to understanding of the activation in the V5 cortical area. On the other hand, knowing what is going on in the V5 cortical area and even following a path from the light of certain wavelengths hitting the retina, through the optic nerve all the way to the integration areas will never tell us what it is like to see colours. This is an example of an epistemic gap.

[23] For the details see the chapter three.

An epistemic gap is simply a situation that we find when considering conscious experiences from the first and from the third person perspectives. In the former case we use concepts that refer directly, i.e. phenomenal concepts, and in the latter case we use concepts that refer via the descriptions of the causal or functional roles. What distinguishes phenomenal concepts from other kinds of concepts that are based on direct reference is the fact that phenomenal concepts use phenomenal states themselves as their mode of presentation. The explanatory gap goes a bit further than that. It claims that there is a gap in our understanding of the epistemic gap. Levine (Levine 2001) argues that there are a few equally legitimate ways of explaining the epistemic gap. According to one of the approaches a difference in modes of presentation reflects a metaphysical difference in properties. The other approach would be to come up with some independent argument for conceptual dualism while simultaneously maintaining ontological monism. The phenomenal concept strategy seems to be doing exactly that. Nonetheless, each one of these approaches to answering the epistemic gap inevitably opens an explanatory gap[24] according to Levine (Levine 2001).

In this chapter I would like to argue that conceivability arguments fail to establish their ontological and epistemological consequences without invoking some additional tacit assumptions. I also argue that the explanatory gap problem seems to arise precisely due to the same intuitions underlying conceivability arguments. One of my main aims in this section is to show that once these additional intuitions are unmasked, the dispute between, on the one side property dualists, epiphenomenalists and mysterians, and on the other inflationist physicalists/conceptual dualists is a matter of a brute disagreement which cannot be resolved by introducing any further arguments that are based on the analysis of meaning. Instead, I will argue that it is a matter of brute disagreement about intuitions that needs to be explained away by some independent accounts such as the quality space model that I discuss in the next chapter.

[24] For the details of this account see the second chapter.

4.2 The ontological gap

Let's start with the ontological gap. Basically, what is claimed by this argument is that if the phenomenal is not logically supervenient on the physical, then phenomenal properties are further properties of the world. Formally the argument goes like this:

20. The fact that conscious experience exists is a positive fact;

21. Conscious experience is not logically supervenient on the physical;

22. If there are positive facts that are not logically supervenient on the physical facts, then physicalism is false;

---

23. Physicalism is false and phenomenal properties are further properties of the world. (Chalmers 1996)

Premise 21. is obviously the most problematic one in the argument[25]. The argument for premise 21. can be expressed as follows:

73. The phenomenal is logically supervenient on the physical

74. If the phenomenal is logically supervenient on the physical, then a situation in which the physical properties occur without the phenomenal properties occurring is not possible.

75. A situation in which the physical properties occur without the phenomenal properties occurring is not possible. (MP 73, 74)

76. A situation in which the physical properties occur without the phenomenal properties occurring is conceivable.

77. If a situation in which the physical properties occur without the phenomenal properties occurring is conceivable, then it is possible.

[25] Much of the current discussions in philosophy of mind focus on this premise.

78. A situation in which the physical properties occur without the phenomenal properties occurring is possible (MP 75, 76)

79. A situation in which the physical properties occur without the phenomenal properties occurring is possible and is not possible (75, 78)

The phenomenal is not logically supervenient on the physical.

The same point can be made in yet another argument and in terms of phenomenal and physical *facts*. It is the famous knowledge argument (Jackson 1982, 1986). As we remember, it relies on the thought experiment involving a brilliant neuroscientist Mary. She is confined to a black and white room and all the resources available to her are only in these two shades. It's important to note that she has a normal vision, the same as anyone who can see the other colours. She is educated and she learns everything there is to know about the neurophysiology, neurophysics, neurobiology, etc. of colour vision. In other words, she knows all the positive empirical facts about colour vision. Once she is released from the room and sees a ripe tomato for the first time, as the argument goes, it seems rather obvious that she learns some new facts that were not contained or inferable from the positive empirical facts that she knew while still in the room. She learns what is it like to have a visual experience of red. Now, the physicalists are committed to the thesis that positive physical facts exhaust all the facts. If physicalism were true she should have known even the phenomenal facts solely on the ground of her knowledge of the physical facts. So the argument formally goes like this:

As we recall from the first chapter, the argument is formalized as follows:

"13. Mary (before her release) knows everything physical there is to know about other people;

14. Mary (before her release) doesn't know everything there is to know about other people (because she *learns* something about them on her release);

Therefore,

15. There are truths about other people (and herself) which escape the physicalist story" (Jackson 1986, p. 293).

Naturally, the premises and the conclusion of this argument have been vigorously debated over the years. Even Jackson himself has changed his mind about the interpretation of the thought experiment. The main point of disagreement concerns whether the knowledge she gains upon release is the knowledge of a further property or whether she learns about an old property under a different mode of presentation.

Regardless of whether someone claims that she doesn't learn anything, or that she learns a new fact but it doesn't reflect a new property, etc., it seems that the way one interprets the argument depends on some tacit assumptions. The matter at hand seems to be a case of a brute disagreement. I will return to this point after recapitulating the main aspects of the explanatory gap problem.

4.3 The underlying intuitions

Conceptual dualists[26] claim that ontologically there is only one property under two distinct modes of presentation. One mode involves descriptions of the causal roles, the other involves direct reference wherein the referent is co-instantiated with the occurrence of a property or its recreation in imagination. Levine asks isn't this arbitrary (Levine 2001, 2006)? He maintains that even if materialism (ontological monism) were true, the other mode of presentation of the same physical property involves some further properties in virtue of which it is different from the physical concept. This in effect does not help the phenomenal concept strategy to completely remove confusion over psychophysical identity. However, as we remember from the section on Max Black's objection, this does not have to be the case. Block (Block 2007) has persuasively shown that identity statements involving

[26] Proponents of the phenomenal concept strategy.

concepts that are based on the different modes of presentation do not necessarily entail metaphysical difference in properties. He argued that it could be that the difference is merely in the two cognitive modes of presentation[27]. Let's go step by step here and see what exactly seems to be the problem and how to remedy it.

Papineau (Papineau 2002, 2007) shows that there is no gap in our understanding of psychophysical identity. Although it is non-explanatory in the sense that genuine identities do not require explanations, once we realize it is a brute identity the confusion vanishes. He admits though that on a first pass there seems to be a gap in our understanding of psychophysical identity, but he explains it away by an appeal to the antipathetic fallacy[28]. In this case the sharper distinction between an epistemic gap and an explanatory gap is really handy. One might say that, according to Papineau's account, there is an epistemic gap but no explanatory gap, or if there is an explanatory gap it is a trivial gap arising from the non-explanatory nature of proper identities. On the other hand, Levine claims that even if he accepts phenomenal concept semantics there seems to be a problem remaining in the intelligibility of explanations. He asks why a certain kind of neuronal activity coincides with what is it like to be in that state, or why should neural states feel the way they do or like anything at all? Papineau concedes on this because of the vague nature of phenomenal concepts[29]. Do these questions make sense? Is it possible to give an intelligible answer to any of them? Perhaps there are two ways of answering questions such as these.

One way would be to say that for example in the natural kinds cases, we can infer or know a posteriori why certain micro properties are accompanied or give rise to particular macro properties because the concepts involved use modes of presentation that are based on the descriptions of the causal/functional roles to connect to their referents. Once we know those descriptions we can work out the details of an explanation. We cannot use the

---

[27] The case of a student of French who thinks he has learned two French words for two different kinds of cat whereas in fact there is only one kind of cat under two different cognitive modes of presentation.

[28] For the details see the third chapter the section of Papineau's account of the phenomenal concepts.

[29] This will be the topic of my whole fifth chapter, so I will leave it here on just a mention.

same method to ascertain how phenomenal states are linked to neural states, if we accept that phenomenal concepts refer directly. However, we might have some other indirect means at our disposal to ascertain the aposterioriti of psychophysical identity. We could say that modes of presentation based on direct reference are compatible with the a posteriori necessity of psychophysical identity. What makes this proposal plausible is that we have very good reasons to believe that physicalism is true, on the grounds of the causal argument[30].

The other way of responding to the objections against the intelligibility of explanation offered by psychophysical identity could be to provide an account of what drives the questions of intelligibility of explanation in the first place. I have in mind some sort of an independent account of the intuition of distinctness inspired by Papineau's (Papineau 2002) notion of " antipathetic fallacy".

In the next section I will work out a bit the notion of the intelligibility of explanation for the ease of explicating my own position.

4.4 The intelligibility of explanations in the new light

Papineau (Papineau 1996, 1998, 2002, 2006) says that it is unintelligible to ask for further explanation of genuine identities, because even in the case of natural kind reductions/identities identity itself is posited as a brute fact and what is further explained is why the instantiation of certain properties satisfies the referent of either a pre-theoretical or a theoretical term. If the explanations involved with the identities were not like this then the explanatory gap would emerge in all kinds of identities and even identities based on descriptions of causal roles.

Levine (Levine 2001) on the other hand says that it is unintelligible how phenomenal and physical concepts could refer to the same properties, however, what seems merely

---

[30] "(1) Conscious mental occurrences have physical effects.

(2) All physical effects are fully caused by purely *physical* histories.

(3) The physical effects of conscious causes aren't always overdetermined by distinct causes.
Therefore, physicalism is true. " (Papineau 2002, pp. 17-18)

unintelligible is why we think that the occurrence of certain neural states is represented via our sensory and psychological concepts.

Levine claims that in the case of phenomenal consciousness there is a problem with the epistemological explanation. This is because an appeal to the brain processes in explaining why I have a particular phenomenal experience of "reddish" quale when looking at a ripe tomato does not seem to explain why I have that particular quale or any other quale for that matter. He goes even further and states that an epistemological explanation of the brain processes does not explain why there should be a reddish quale that there is. (Levine 2001, p. 7)[31].

Furthermore, if they were indeed metaphysically distinct then it would have perhaps been unintelligible why they are identical. Given that they are one and the same thing then it doesn't make sense to ask this question. Why does Levine not accept that the same property can be picked out in two different ways? The matter again concerns/relates to intuitions and brute disagreement. Levine doesn't conclude they are metaphysically distinct, instead he seems to start with that intuition.

Levine (1983, p 357) claims that psychophysical identity leaves an explanatory gap that doesn't seem to emerge in the case of natural kinds. He says that we don't encounter the explanatory gap with natural kinds because the explanation is fully intelligible as it is given in terms of the descriptions of the causal roles properties standardly play. To put it differently, the explanation given in these terms exhausts all there is to/to know about the notion of a given natural kind. This is not the case with qualia, because the concept of pain, if based on a description of causal roles, does not exhaust all there is to/to know about pain, thus opening the explanatory gap. This is because there is something about pain that is knowable only from the first person perspective and as such in principle cannot be derived from or contained in a description of causal roles. Contrary to Levine's claims, it has been argued in the preceding chapters that the causal roles descriptions do not in fact exhaust all there is about a given property but in fact help us understand why certain concepts satisfy

[31] For the difference between metaphysical and epistemological modes of presentation see the second chapter.

the descriptions associated with the instantiation of the respective properties. Now, Levine argues that an explanation that is based on descriptions of the causal roles leaves out the what-is-it-like element of the experience thus opening the explanatory gap. I agree that explanations of phenomenal consciousness involving descriptions of the causal roles may seem explanatory inadequate. However, as we have seen in the section on the inconsistency objection to the knowledge argument, if phenomenal properties were causally inefficacious then we wouldn't know about them from either the first or the third person perspective. So this feature of phenomenal knowledge is perfectly utilized in the phenomenal concept strategy and perhaps stands as an independent argument in favor of this strategy.

Let's get back to the main line of argumentation. Levine writes in his (Levine 1983, p. 359) that there is nothing in an explanation of C-fiber firing that tells us why it feels the way it does. If that is the case then it becomes conceivable straight off that C-fiber firing could occur without anything it is like to be in that state. The physicalist would not want to embrace this consequence because, if physicalism is true, then it has to be necessarily true, so the counterfactual claim of their coming apart should not be possible. So he departs from Kripke's original metaphysical point and claims that it's possible that we could have a characterization of qualia that would make it equally incoherent to imagine that C-fiber firing occurs without pain quale, thus blocking the entailment from conceivability to metaphysical possibility. However, the claim that there is nothing in the description of "C-fiber firing" that determines or explains why it feels that way, amounts to saying that what it is like to be in pain is not exhaustively analyzable in terms of the descriptions of the causal roles of the underlying physical processes. Of course, "the C fiber firing" does not capture "what it is like to be in pain" because the latter we know in a substantial, determinate and direct way, and it can only be known in that way, whereas to know the former only descriptions of the causal roles would suffice.

As Papineau (Papineau 2007) rightly notes, the often used examples of the natural kinds identities in the anti-physicalist claims, are not in fact examples of the scientific explanations of identities. That is, these explanations do not amount to explaining why, say, water is H2O, it would not make much sense. They rather serve to explain why water has some further properties like being colorless, odorless, tasteless, etc. We are explaining why the pre-theoretical notion satisfies the causal roles descriptions to which we associate it. So,

even in the natural kinds cases the descriptions of the causal roles do not explain the identity itself. If we accept that phenomenal concepts refer directly, not via descriptions of their causal roles, then why would we concede on the requirement that the brute identity of the phenomenal/neural have a further explanation either? Given that proper identities are brute identities that do not require further explanation, and given that psychophysical identity is one of them, it actually doesn't make it intelligible to ask for further explanations.

If neurobiological explanations that use concepts that refer via descriptions of the causal roles, do not make it intelligible why, say, C-fiber firing feels the way it does and if phenomenal concepts, on the other hand, explain how the same property is picked out by two different concepts, then what explains why the C-fiber firing feels the way it does? One way to tackle this issue is Papineau's explanation of the intuition of distinctness in terms of antipathetetic fallacy[32].

Faced with this dilemma we can argue that, although there is an epistemic gap between natural kinds and phenomenal kinds, there is no gap in our understanding of their identity, even though brute identities are in a sense non-explanatory.

Papineau persuasively shows that psychophysical identity is inexplicable as a proper kind of identity. However, he concedes that the epistemic gap exists. This situation I think fits perfectly with the sharper distinction between the epistemic gap and the explanatory gap mentioned earlier. Phenomenal concepts refer directly not via causal roles descriptions, this makes psychophysical identity as inexplicable as the other identities in science (because in those cases we are not explaining the identity via causal roles, but why concepts satisfy causal roles of the further properties).

We mentioned at the beginning of this section that the explanatory gap problem has three main aspects: one is conceptual-given that natural kind concepts refer via descriptions of their causal roles it's not clear how phenomenal concepts would pick out physical properties; the second aspect is modal-given that the phenomenal is not explicable in terms of causal roles, then the contingency of the psychophysical identity seems, at the very least,

[32] For the details see the chapter three.

epistemically possible; and the third is that, given the semantics of phenomenal concepts, is it still intelligible to claim that there is a gap in our understanding of psychophysical identity.

Papineau's theory answers very plausibly to all three aspects, but would it be possible to come up with an independent analysis of the third aspect? Namely, in his arguments he combines the semantics of the phenomenal concepts based on direct reference with explanatory asymmetry in order to reject a priori characterization of physicalism, i.e. the characterization according to which phenomenal facts should be a priori derivable from the complete physical facts. In the following sections I shall be examining a brute disagreement of intuitions underlying the explanatory gap. In the final section I will provide another take on intelligibility of explanations and a priori derivation.

4.5 The brute disagreement

Finally, we have come to the point where central problem can be articulated. Let me sketch two main aspects of the problem.

A) Firstly, it is not obvious nor prima facie plausible to assume that the two ways in which we conceptually grasp physical and phenomenal properties, the former by descriptions of their causal roles and the latter by direct reference, reflect two metaphysically distinct properties. Standardly, arguments that corroborate this claim rely on the intuitions one starts with and some additional assumptions. The really problematic intuitions are the starting intuitions. Even the language used to set up the problem is dualistic. For example: how do physical properties ***give rise*** to phenomenal properties, etc. If they are the same then one doesn't give rise to the other. They are one and the same thing. The mere fact that we have a different cognitive access to the brain processes and to "what it is like" to undergo them does not itself make it obvious that there are two metaphysically distinct things. There are no a priori reasons to assume that they are distinct. When the matter at hand is about fundamental intuitions, then no further argument would help to resolve the issue. This is the case with a brute disagreement. What is instead needed is an account that would explain away those intuitions.

B) The second issue is a matter of the intelligibility of explanations. It is usually argued that, even if phenomenal properties were in fact physical properties but picked out in different ways, it would still remain puzzling why certain neurobiological states are accompanied by particular phenomenal states or any other states at all for that matter. However, if there is no independent argument against the claim that there being two distinct ways of picking out the same property is compatible with truth of physicalism, then again this issue reflects the same case of a brute disagreement in tacit assumptions. The same point about different modes of presentation applies to the issue of explanation as well. Instead of trying to come up with an explanation of the explanatory gap in terms of a priori derivation from basal truths, an explanation that is itself based on descriptions of causal or functional roles, I suggest we look for another independent explanation of the apparently disrupted link between conceivability and explanation.

Let me give an illustration of what I mean by the brute disagreement.

4.6 The tacit assumptions

Even if thought experiments in philosophy of mind had bearing on our understanding of psychophysical identity it would not be because certain intuitions that drive the experiment turned out to be true, but because of the tacit additional assumptions involved. If we vary these tacit assumptions and keep the core of a thought experiment constant, we get radically different conclusions. The question arises, what is the use of thought experiments and intuitions if they themselves don't change the conclusion of the argument? We need to rethink the whole issue.

Take for example the knowledge argument (KA). What follows from the KA alone depends on some additional assumptions. Let's have another look at the original Knowledge Argument. As we recall it goes like this:

"13. Mary (before her release) knows everything physical there is to know about other people;

14. Mary (before her release) doesn't know everything there is to know about other people (because she *learns* something about them on her release);

Therefore,

15. There are truths about other people (and herself) which escape the physicalist story" (Jackson 1986, p. 293).

The conclusion in 15. does not follow merely from the assumptions that she knows all the physical facts and that she learns something new upon release. The argument itself does not warrant that conclusion without the additional assumptions about modes of presentation.

It seems that the argument profits on different readings of the notions "knowing everything physical" and "learning something new". Namely, Jackson seems to tacitly assume that concepts which we use to infer new facts from the physical facts can only rely on descriptions of the causal roles as modes of presentation. Perhaps that might be true if one wanted to infer new facts directly from the physical facts solely on the ground of descriptions of the causal roles. However, a mere fact that Mary learns something new does not justify an assumption that what she learns about escapes physicalists story. It could be the case that what she learns from the first person perspective is not directly inferable from the body of physical information, because, as we saw in the chapter on the phenomenal concepts, the concepts we use to think about our experiences from the person perspective refer directly, and cannot have a priori connections with the material properties, thus cannot be inferred from a scheme of causal roles anyway. I would be inclined to say that not much in the argument depends on the issue of whether she learns something new, but rather depends on some assumptions about a relation between conceptual semantics and a theory of explanation.

For example, if we grant that Mary learns something new and characterize physicalism as an identity thesis that is based on distinct conceptual semantics, on the one hand based on direct reference and on the other based on the descriptions of the causal roles, then it is reasonable to conclude that conceptual dualism (ontological monism) or representationalism are true. On the other hand a characterization of physicalism, which

employs conceptual semantics based only on reference via causal roles, gives us property dualism or epiphenomenalism.

This argument seems to establish merely epistemic gap, i.e., given the premises alone one can only conclude that there are two distinct ways of knowing something. A conclusion that there are truths about Mary and other people which escape the physicalist story requires an additional assumption according to which physicalists have to rely only on concepts that use descriptions of the causal roles as modes of presentation. However, that assumption is not warranted merely by the fact that she learns something new.

As we can see, when the core of the knowledge argument remained the same whereas assumptions about the modes of presentation were varied we got significantly different conclusions. The question immediately arises: do we have any reason to prefer one set over the other set of the additional assumptions? Perhaps we do.

The point is that the KA can support all positions: conceptual dualism and epiphenomenalism and property dualism. We have seen that what we conclude from these arguments depends strongly on the assumptions we start with. One might of course say that we use philosophical arguments to test our intuitions, but as my analysis shows, the same argument can support even opposed positions. We can of course come up with some even more ingenious arguments to test our intuitions, but no arguments would resolve a clash of brute intuitions. The question that I would like to raise here is: do these intuitions encapsulate theoretical commitments rather than serving as an independent support for the claims in the arguments? I would be inclined to say that the latter is probably true. As David Rosenthal puts it:

“Perhaps the most significant contribution of experimental philosophy, then, is to underscore the likelihood that various intuitions are not theory neutral, but products of such theoretical commitments. If intuitions aren't widely shared by the folk and fit conveniently with theories espoused by those who invoke the intuitions, it's natural to ask whether the intuitions encapsulate those theories, rather than providing independent support” (Rosenthal 2010, p. 5).

In my opinion, what is needed instead is an account that would explain away the underlying intuitions. Chalmers’s account is that the epistemic gap reflects or is based on the

ontological gap; Papineau had offered an interesting account for the antimaterialst's intuition of distinctness (Papineau 2002)[33]. That is the intuition that the two ways we think about consciousness reflect two metaphysically different properties. He calls it the intuition of distinctness. Perhaps in this case we could come up with an independent account for the intuition of distinctness. The central strategy should then move from arguments for certain positions to accounts that explain away the intuitions that underlie those arguments.

As we remember, according to Levine (Levine 2001) the explanatory gap problem arises because we don't know why certain brain states are accompanied by particular phenomenal states. It is supposed to be a mystery. However, once we realize that we don't know the causal roles of phenomenal properties based on knowledge of physical properties because phenomenal properties don't refer via descriptions of causal roles in the first place, it becomes clear that the only remaining job is to explain away the "intuition of distinctness", i.e. to explain away why we assume or have a feeling that phenomenal concepts ought to refer via descriptions of causal roles.

Even the question whether Mary learns something new does not follow from the argument itself or from the intuitions alone. It depends on tacit additional assumptions about modes of presentation. If we assume that the way the phenomenal concepts pick out their referents should be based on descriptions of their causal roles, then two avenues are open. On the one hand, she should not have learned anything new because it is possible to derive or to infer the phenomenal facts from the physical facts only if we had complete enough information about physical facts. On the other hand, she could have learned something new, but in this case what she learned is not physical information, because if it were she should have been able to infer it from the basal physical facts and descriptions of the causal roles. It is worth noting that the crucial step from learning something new to learning about a new metaphysically distinct property would require an additional argument that is not provided in the Knowledge Argument.

[33] I have discussed this idea in part in the first chapter and at greater length in the previous chapter. For the details see the section on Papineau's account of phenomenal concepts.

However, if we assume that the phenomenal concepts refer directly, not via descriptions of their causal roles, we get a totally different story. On this account, Mary learns something new that was not inferable from the physical facts simply because phenomenal concepts do not use descriptions of their causal roles as modes of presentation, and thus cannot be inferred in that way.

Now, the first option does not allow for differentiating between an eliminativists[34] interpretation on the one side and the epiphenomenlist/property dualist[35] interpretation on the other. On the other hand, the conceptual dualist option seem to address very well both why phenomenal concepts are not contained in nor inferable from the body of physical information and how their existence is compatible with physicalism. One could stop right here and say that, for example, the explanatory principle of "the inference to the best explanation" or Ockham's razor suggests which one of the two interpretations should be preferred. Of course it is the conceptual dualist's interpretation. But we might go even further and consider the reasons why the property dualist interpretation should be favoured, put forward by Chalmers and Jackson (Chalmers and Jackson 2001). They believe that there has to be some a priori element in the reductive explanation on which we base the necessity of even the a posteriori identities. As we shall see, no such analysis should be required for the psychophysical identity as it is not universally required for some of the standard cases of reductive explanation of natural kinds (Block and Stalnaker 1999).

An argument for conceptual dualism then, might stem from considerations of conceptual analysis and intelligibility of explanation. I have in mind Block and Stalnaker's analysis of the account of reductive explanation given by Chalmers and others (Chalmers 1996; Chalmers and Jackson 2001; Levine 2001). They assume that identifying explanans and explanandum via conceptual analysis is universally applicable to all cases of the reductive explanation of natural kinds, but it is not available for psychophysical identity. Block and Stalnaker persuasively demonstrate that this is not the case with many examples of the reductive explanation of natural kinds. Furthermore, in these cases we find successful

[34] Eliminitavists maintain that there is no such thing as qualia and that folk psychology miss-describes experiences.

[35] Chalmers is the main proponent of this position as it was discussed in the first chapter.

reductive explanation without relying on conceptual analysis. If that is the case then why would conceptual analysis be required for psychophysical identity? It seems as if our preferences over the tacit assumptions involved in the brute disagreement we mentioned earlier should swing us in favour of the phenomenal concept strategy. I need to elaborate on these claims before concluding my key argument.

4.7 A priori derivation and the intelligibility of explanations

Block and Stalnaker (Block and Stalnaker 1999) argue that there are two main claims associated with the problem of consciousness. One is the claim about the explanatory gap problem, i.e. that any physicalist or functionalist notion of consciousness leaves out the fundamental feature of consciousness, the what-it-is-likeness of experience. According to some philosophers this gap is closable if we acquire relevant concepts (Nagel 1974), others claim that the gap is not closable because it reflects a fundamental part of consciousness which is not physically or functionally analyzable (Chalmers 1996; Chalmers and Jackson 2000; Jackson 1982) because of its non-physical nature. The concept of consciousness is not functionally analyzable therefore zombies, creatures that are our physical and functional isomorphs but who lack phenomenal consciousness, are metaphysically possible. The other claim about consciousness is that conceptual analysis of consciousness is not available. Conceptual analysis in this case is to be understood as an analysis according to which phenomenal concepts are a priori derivable from the complete microphysical description.

These two claims are mutually dependent. Specifically, the lack of availability of a conceptual analysis of consciousness stems from the fundamental nature of consciousness, this sets it apart from other physical phenomena. It should be fairly uncontroversial to say that, if consciousness were functionally analyzable, there would be no explanatory gap; psychophysical identity would be considered as any other identity of natural kinds. The controversial issue according to Block and Stalnaker (Block and Stalnaker 1999, p. 2) is whether such an analysis is required to close the gap. According to some philosophers, in particular Chalmers and Jackson (Chalmers and Jackson 2001), such an analysis is not

possible to give therefore the explanatory gap cannot be closed. Let's follow Block and Stalnaker and call this *the conceptual analysis thesis*.

When it comes to explaining natural kind phenomena Levine (Levine 1993)[36] maintains that it involves basically two phases. One phase consists of working out the causal or functional roles, and once this is done one can a priori identify explanans and explanandum. The other step is an empirical one where one actually has to find out empirically what the causal or functional roles are, or finding the fillers of these roles. The problem of course, according to Levine, is that no such analysis is available for phenomenal consciousness. I might add that this is because phenomenal concepts pick out their referents directly, so knowledge of their causal or functional roles does not figure anywhere in the workings of the phenomenal concepts[37]. In other words, the first step of the Levine's two-step program is totally missing in this case. We can't analyze phenomenal concepts in terms of causal roles because they do not pick out their referents via descriptions of causal roles. As we shall see in the next chapter when discussing the vagueness of phenomenal concepts, this should not pose a problem for physicalism as it only means that the intelligibility of explanation regarding psychophysical identity is to be sought elsewhere, not in conceptual semantics.

As it was said, the really controversial issue here is whether such a conceptual analysis is always available even for the natural kinds cases. In another words, Block and Stalnaker (Block and Stalnaker 1999, p 14) hold that such an analysis is in fact not always available for other successful reductive explanations of natural kinds. Block and Stalnaker (Block and Stalnaker 1999) cast serious doubts on the claim that closing the explanatory gap would require a priori conceptual analysis as proposed by Levine (Levine 1993), Chalmers (Chalmers 1996) and Chalmers and Jackson (Chalmers and Jackson 2000). They argue that Levine's and other's arguments only show that such an analysis might be required but not that it must be required. Basically, Chalmers and Jackson use only examples that already involve a priori conceptual analysis but they are silent on other cases of natural kinds that

[36] (Chalmers 1996) and (Chalmers and Jackson 2000) position themselves along these lines of thinking as well.

[37] For the details see the second and the third chapters.

are based on successful reductive explanations that do not involve a priori conceptual analysis. Block and Stalnaker consider examples of natural kinds in which the explanatory gap was closed but no a priori conceptual analysis was available. Such an example is the "explanatory gap" in explaining life (Block and Stalnaker 1999, pp. 14-16). In explaining what life is we analyze the concept of "life" in terms of functional roles such as digestion, reproduction, locomotion, respiration and so forth. However, it might have happened that some living organisms never digested, respired, reproduced or moved, some tree-like creatures perhaps, but were still alive. Perhaps they used some alternative processes instead of digestion, reproduction, locomotion, respiration, etc to survive. This example shows that the concept of life is not *a priori* analyzable in terms of digestion, reproduction, locomotion, respiration, etc. In effect, it shows that to close the explanatory gap on life no a priori conceptual analysis is required. As the authors put it: "Closing the explanatory gap in the case of life has nothing to do with any analytic definition of 'life', but rather is a matter of showing how living things around here work" (Block and Stalnaker 1999, p. 15).

Basically, Jackson and Chalmers believe that a priori conceptual analysis can take two forms: one form of analysis is in terms of derivation from the microphysical truths, the other form is in terms of reference fixing definitions, like the well known cases of "water" or "heat". On the other hand, Block and Stalnaker argue that neither of these two versions of conceptual analyses apply universally to all varieties of natural kind reductive explanations.

The problem with Jackson and Chalmers' view concerning reductive explanations and a priori analysis is that they seem to believe that the way in which we connect a microphysical description, such as "H_2O" or "molecular kinetic energy", with the concepts of macro phenomena, such as " water" or " heat" is through conceptual analysis. If an explanation of identity statements relied solely on the descriptions of causal roles it would merely amount to establishing a correlation. However, properly understood, identity allows for transferring explanatory and causal force to the new facts, this is not the case with mere correlations. So it is legitimate to ask a question why certain correlations hold, e.g. if the relation between "H_2O" and "water" were a mere correlation and not a proper identity, then it would make sense to ask why they are correlated, etc. However, if they were identical there is nothing more to it that needs to be explained. Consider this example. Imagine there were two independent literary societies, one studying the work of Samuel Clemens and the

other studying the work of Mark Twain. By coincidence, both societies organize an annual meeting in the same hotel and at the same time. At some point during the dinner in the restaurant both groups chose to gather after the conference and someone from the "Clemens" group drops a document that describes the life and work of Samuel Clemens. Someone from the "Twain" group picks it up and astonished says: "what an amazing correlation, the life and work of Samuel Clemens is the same as of Mark Twain!" Note that in this case it makes perfect sense to ask for an explanation of a correlation. However, once it is ascertained that "Mark Twain=Samuel Clemens" it doesn't make sense to ask for any further explanations, because identities don't have explanations (Block and Stalnaker 1999, p. 24).

The examination of the claim that a proper reductive explanation requires conceptual analysis in terms of a priori derivation from microphysics shows that, even if there were some a priori analysis of the macro-concepts available, the whole strategy would fail because the analysis would not be made in terms of microphysical claims (Block and Stalnaker 1999, p. 28).

Consider this argument pattern as an illustration:

"78. 60 percent of the earth is covered with H_2O.

79. H_2O is the stuff that plays the right kind of causal role in explaining our use of the word 'water'.

80. Therefore, 60 percent of the earth is covered with water.

The assumption must be that the inference from 78. and 79. to 80. is now mediated by the following conceptual analysis:

81. Water is the stuff that plays the right kind of causal role in explaining our use of the word 'water'." (Block and Stalnaker 1999, pp. 25-26).

The crucial step in the argument that is supposed to grant that 80. follows a priori from the premises 78. and 79. is of course hidden premise 81. However, 81. is not a conceptual truth about the meaning of the word “water”. At best it is an abbreviation of an a posteriori and empirical analysis of the right kind of causal role of water, but their account provides no reason to assume that eventually we will be able to fill in the appropriate a priori analysis of the term “water”. For all we know some other liquid might have played that role. On the other hand, granting for a moment that an a priori conceptual analysis can be found in a premise like 81., then there is nothing that would prevent the same kind of analysis pattern for psychophysical identity.

If this kind of conceptual analysis is available for natural kinds like in the first analysis pattern then it is also available for cases involving phenomenal concepts. Furthermore, it certainly does not constitute a difference between the conceptual analysis of natural kinds and the analysis of phenomena concepts.

Consider this example:

82. C-fibre firing is taking place in Jones at time t.

83. C-fibre firing is the process that plays the right kind of causal role in explaining our use of the word 'pain'.

84. Therefore, pain is taking place in Jones at time t.

The inference from 82. and 83. to 84. is justified by the following quite trivial instantiation of our account of reference:

85. Pain is the process that plays the right kind of causal role in our use of the word 'pain' (Block and Stalnaker 1999, p. 26).

The Block and Stalnaker examination of the account of conceptual analysis and reductive explanation proposed by Levine, Chalmers and Jackson shows that possibility of zombies poses no threat to physicalism because no a priori analysis in terms of microphysical descriptions is required to reductively explain mental phenomena. That is to say, either the meaning of concepts is not a priori analyzable or it is not analyzable in terms of microphysical descriptions. Remember, zombies posed a threat to physicalism in the first place only because it was argued[38] that even a posteriori necessary statements require a priori analysis of the meanings of their concepts at some stage. So it seems that the only way to save the conceptual analysis strategy is by bypassing the issue of a priori analysis in terms of microphysical descriptions and trying instead to find an a priori analysis of meaning in some other terms.

This point also applies to the issue of the intelligibility of explanations because it seems that Chalmers and Jackson hold that an intelligible reductive explanation must follow a priori from the microphysical description. This assumption is actually the main motivation for their move from the epistemological plain to the metaphysical one in the conceivability argument. Namely, that what is merely conceivable is metaphysically possible.

Moving on. In order to justify conceptual analysis, Chalmers and Jackson (Chalmers and Jackson 2000) appeal to analysis through the two-dimensional framework by claiming that, instead of the a priori analysis of macro-physical concepts in terms of microphysical descriptions, the concepts get their meanings through the special nature of primary and secondary intensions[39]. As we shall see, this move does not help to establish that a priori conceptual analysis is required for natural kinds concepts or phenomenal concepts either.

As Block and Stalnaker put it:

"The two-dimensional account does nothing at all to motivate the claim that there is an a priori accessible conceptual component of content" (Block and Stalnaker 1999, p 30).

---

[38] For the details see chapter one section on conceivability arguments and chapter two section on intelligibility of explanations.

[39] For the details of the two-dimensional framework see the first chapter section on the 2D argument.

As we remember from the section on the two-dimensional argument, Chalmers provides an interpretation of two-dimensional semantics that is based on the distinction between primary and secondary intensions of meaning. A secondary intension is a function that reflects how the reference is fixed in a counterfactual world considered as actual (centred). We know it from the Kripkean cases: a secondary intension of "water" in the actual (centred) world is "H_2O", whereas at Twin Earth considered as actual (centred) it is "XYZ". A primary intension on the other hand is a function that is determined by how the reference would have been fixed in a given world considered as actual (Block and Stalnaker 1999, p. 33). One can know the primary intension of a given concept and still be ignorant of the secondary intension, because the secondary intension varies from possible world to possible world, note, secondary intension is a function of how the reference is fixed in a possible world considered as actual (centred). In this sense then, the primary intensions should represent the a priori part of the concept's content that remains unchanged when considered through possible worlds. The primary intension understood in this way represents the description of the causal roles that played the central part in the discussion on conceptual analysis, and thus can be understood as representing the a priori part of the two-dimensional analysis.

We can see now how the two-dimensional framework is supposed to answer the worries that conceptual analysis generated. In particular, worries about how a microphysical description enables a priori inference to macro phenomenal concepts through descriptions of the causal roles.

Consider a familiar example wherein X is a primary intension that is known a priori:

86. 60 percent of the globe is covered by H_2O.

87. H_2O= the satisfier of X (the primary intension of 'water').

88. Water = the satisfier of X.

89. Therefore, 60 percent of the globe is covered by water. (Block and Stalnaker 1999, p. 37)

For this argument to work 87. has to be a microphysical truth and 88. has to be a conceptual truth. In other words, 87. is supposed to be an empirical claim whereas 88. should be a priori inferable from 87. However, the argument does not warrant that, at least unless we assume that all the facts are derivable or contained in the microphysical facts. But this claim seems pretty debatable. For consider the following case. Suppose there are two possible worlds that are microphysically indiscernible, call them W-earth and W-super earth. On what ground can we suppose that a microphysical description of W-earth describes only and uniquely the W-earth? Given that primary intensions are just functions it could happen that, given different inputs to otherwise indiscernible possible worlds, the primary intension of water at W-earth leads to H_2O but at W-super earth the primary intension leads to something else. If this is correct then the primary intension of water although a fact, would not be a microphysical fact. If this is the case then we have disrupted the inference between microphysical descriptions to a priori analysis thus begging the question why the same pattern of analysis can't be applied to the case that involves C-fibre firing and pain? It seems that there is nothing that prevents us from applying the same pattern of analysis as long as we stick to the interpretation of the two-dimensional semantics according to which the primary intensions are derivative from the actual and possible secondary intensions at any given possible world (Block and Stalnaker 1999, p. 41).

Here is how Block and Stalnaker put this point:

"90. Pyramidal cell activity was rampant in medieval prisons.

91. Pyramidal cell activity = the satisfier of the primary intension of 'pain'.

92. Pain = the satisfier of the primary intension of 'pain'.

Therefore,

93. Pain was rampant in medieval prisons.

Our point is that there could be compelling motivation for 91. and that 92. has

whatever a priori status 88. above has." (Block and Stalnaker 1999, p. 44).

This sort of analysis shows that the two-dimensional framework does not provide an account of how to isolate the a priori component of the reductive explanation. It rather seems as if it is merely assumed. Together with the examination of the conceptual analysis thesis this shows that the reductive explanation of natural kinds does not universally require a priori derivation of macro-physical concepts from microphysical descriptions. This claim is based on the consideration of cases of conceptual analysis in which either a macro-physical concept is not inferred from the microphysical description or if it is based on the microphysical description then the inference is not a priori. Similar points can be applied for the two-dimensional framework which tries to evade the just mentioned difficulties of conceptual analysis by introducing a primary intension of meaning that captures the a priori part of the descriptions of causal roles. If the primary intension is determined by the secondary intension, then in this case there is no reason to believe that is the case merely on the grounds of the microphysical descriptions.

If neither of these methods provide an account that is universally applicable to the natural kinds, then there is no reason to believe that it could not work in the case of psychophysical identity. My point here then is if proper identities don't require further explanations, and if we use identities to explain some further properties, i.e. to transfer the explanatory and causal force onto these further properties, then a justification of the intelligibility of such an explanation should be sought somewhere aside of in the a priori derivation from microphysical descriptions. The real justification of the intelligibility of a reductive explanation of phenomenal consciousness is to be found in the issue of why thinking that intelligible explanation requires employment of modes of presentation that are based on descriptions of causal roles. As we have just seen, if a justification along the lines of a priori derivation from microphysical descriptions, with or without isolating the a priori part of meaning through the two-dimensional framework/through two-dimensional frameworks, fails, then it should be found somewhere else. I suggest that the right place to start looking is the structure of relations among sensory concepts, qualitative concepts, phenomenal concepts, neurobiological concepts, neuroanatomical concepts, and psychological concepts, etc. On my proposal this structure of relations can be outlined by using a quality space model. More about that will be said in the next chapter.

My agenda in this chapter was to show that metaphysical modes of presentation play no role in the intelligibility of explanations that involve phenomenal concepts. I would rather claim that there is no reason to suppose that a merely epistemic difference between the concept of experience from the first person perspective and the neurobiological concept of experience can be explained away solely by metaphysical modes of presentation. If such an explanation is not based exclusively on metaphysical modes of presentation, then there is no prima facie reason not to accept conceptual dualism while maintaining ontological monism. I tried to demonstrate that there are no a priori reasons to accept the claim that the a priori derivation of phenomenal concepts from the microphysical description should be required for an intelligible explanation of psychophysical identity. Thus I argued that the only remaining thing that needs to be explained is the intuition of distinctness and why conceptual dualism stands better off. For that I introduced my "brute disagreement" account. To explain away this brute disagreement I propose to use a quality space model that would tease out a structure of relations among the diverse kinds of concepts that are used in the study of phenomenal consciousness. I will be discussing the quality space model in much greater detail in the next chapter.

5. CHAPTER 5

5.1 Intuitions of distinctness and quality space

"However much we know about our cerebral innards, and however varied the examples of human pain we are given, there will still be a number of distinct material properties which this sort of research will be unable to decide between as the material essence of pain. "

(Papineau 2002, p 177)

5.2 The vagueness of phenomenal concepts as a constraint for studying phenomenal consciousness empirically

As it has been established so far in our discussion, an a posteriori link between a material property and the concept that picks it out is to be determined empirically. If we accept that physicalism can be set up as a psychophysical identity thesis, that is as an a posteriori necessary thesis, and, given that phenomenal concepts have no a priori links with physical properties, it is difficult to determine which particular material property is picked out by any particular phenomenal concept. That is to say, if physicalism were true then each phenomenal concept ought to refer to some particular material property, however, given that phenomenal concepts do not use descriptions of causal roles to pick out their referents it is not clear how to set up an empirical enquiry about which phenomenal concept picks out which particular material property.

We could for example try identifying the material referent of a phenomenal concept, say of the concept “pain”, by mirroring introspective reports of each case of “pain” individually with the particular phenomenal concept, and creating a comprehensive database for such cases. However, regardless of how labour intensive and unpromising this approach may seem, there is a more obvious shortcoming built into it. Namely, this approach would not allow us to generalize over other animals, robots or extraterrestrials, unless of course, we examine each of them individually. This is not merely a technical or methodological obstacle that can be eventually solved, but rather a principled issue that stems from the vagueness of the phenomenal concepts themselves and is in principle unsolvable. As we remember from the previous chapter, the Knowledge Argument established that what Mary learns upon her release is not based on descriptions of causal roles, for if it were, she should have been able to know it while still in the room. On the other hand, she could sit in front of a pile of ripe red tomatoes as much as she likes and introspect her experiences of the red colour, but that would never enable her to infer the descriptions of the causal roles involved. On the other hand, descriptions of causal roles would never tell us what it is like to feel pain, because knowing that requires actually undergoing such an experience. Phenomenal concepts refer directly and her knowledge about what she introspects is substantive and determinate. Although our knowledge of the experience from the first person perspective is substantive and determinate, phenomenal concepts are vague, because, as it has been said, they have no a priori connections with their material properties. This is the reason why direct mapping between phenomenal concepts and material properties is not possible.

Direct mapping of introspective reports onto the MRI or PET scans on the other hand can be extremely useful for enhancing a scheme of the descriptions of the causal roles. That is to say, it can be very useful for working out the psychological concepts involved. Psychological concepts are very important for finding material properties corresponding to appropriate phenomenal concepts because they furnish the database with the descriptions of the causal roles that we need to get the research off the ground in the first place.

The relation between phenomenal concepts and the explanatory gap is complex. On the one hand, because the phenomenal concepts are vague, qualia inversions are

conceivable but not metaphysically possible[40]. On the other hand, due to the phenomenal concepts' vagueness we can't know why certain phenomenal states are accompanied by certain brain states, remember they have no a priori connections with the material properties and they do not refer using descriptions of causal roles. But, because of the phenomenal concepts' vagueness, the explanation as to why certain phenomenal properties are correlated with certain neural or physical properties should not stem directly from the phenomenal concepts themselves, as it does in the case of natural kinds concepts. As we remember, in the case of natural kinds the concepts use descriptions of causal roles, and so, based on these descriptions, we can in principle know which concept is associated with which property. Since we don't have anything like that in the case of phenomenal concepts because they refer directly, unmediated, an account that would explain which phenomenal concept is associated with which property cannot stem directly from the conceptual semantics as with the case of natural kinds. Obviously we need an account independent from the phenomenal concepts' semantics to explain which phenomenal concept is associated with which material property. I suggest using a quality space model as a tool for analysis of different psychological concepts in this area that would helps us better understand why phenomenal concepts are expected to refer via descriptions of causal roles. This expectation is in my understanding of what Papineau (Papineau 2002) calls "the intuition of distinctness".

The way in which Papineau portrays the intuition of distinctness inclines towards a sociological/anthropological explanation of it encapsulated in his famous "anti-pathetic fallacy"[41] (Papineau 1993). The reason for introducing the anti-pathetic fallacy comes mainly from the causal argument for physicalism (Papineau 2002). Papineau argues that there can't be any other reason for not accepting conceptual dualism, given the causal argument, except a mere intuition deeply rooted in our thinking about consciousness.

As we remember, Papineau's causal argument is a really powerful argument, however, it only gives us really strong reasons to hold that physicalsim is true. But it does

[40] For the details see chapter two, section on the non-ascriptivism.

[41] For the details see the chapter three.

not follow directly from the argument that conceptual dualism is true[42]. For that we need an independent argument or account. Conceptual dualism is one of the ways of explaining the a posteriori necessary nature of psychophysical identity. So although we have very good reasons to believe that physicalism is true, we nevertheless need an argument for conceptual dualism. The issue at stake concerns the vagueness of phenomenal concepts. That is, the truth of the conceptual dualism implies vagueness of phenomenal concepts, which in effect amounts to saying that we cannot know precisely which physical property is associated with which phenomenal concept.

This, however, is a different issue from what Levine calls the issue of the unintelligibility of explanation. It has nothing to do with the claim that we do not know why certain brain states are followed by certain mental states, or any states at all. According to the phenomenal concepts strategy, such an answer is impossible to give in principle because phenomenal concepts don't refer via descriptions of causal roles as modes of presentation. On this understanding, according to the phenomenal concept strategy, we will never be able to pinpoint the exact "neural correlate of consciousness" directly from the phenomenal concepts. But this claim only amounts to saying that an account of which phenomenal concept is connected to which material property is to be sought somewhere other than in the semantics of the phenomenal concepts directly, i.e. such an account should be external to the phenomenal concepts themselves.

Perhaps there is another way to answer this worry. A way that is different from the model of conceptual analysis or a priori derivation from microphysical descriptions, proposed by Levine, Jackson and Chalmers[43]. Perhaps an examination of the structure of relations among concepts in the relevant areas of science that are trying to capture the phenomenal character would point towards a criterion for which theories are better suited to the task.

---

[42] It only follows that phenomenal and physical causes are identical.

[43] For the details see the last section of the previous chapter.

Positing phenomenal concepts is supposed to secure the a posteriori necessity of psychophysical identity and, given the built-in vague nature of phenomenal concepts, this cannot tell us why certain physical properties are identical with particular phenomenal properties. Identities don't do that kind of explanatory work. What the identity does is transfer explanatory and causal force onto some further properties (Block and Stalanker 1999). On the other hand, the quality space model that does not directly map qualitative character/phenomenal properties onto perceptible properties/physical properties, but rather allows for the mapping of the structures of resemblance between two or more properties of multiple sensory modalities and further among multiple conceptions that capture those modalities, can help us distinguish why phenomenal concepts are expected to refer via descriptions of their causal roles in order to be fully explanatory. On my proposal, given the difficulty of constructing a straightforward quality space for particular modalities of sensory experience, such as pain, there could be different levels of/within quality space, there could be multiple quality space axes within a single quality space axis so we could zoom into and zoom out of the structure of their functional roles. Before working out the details of such a quality space I need to examine the relation between the intuitions of distinctness and the vagueness constraint, which I do in the next section.

5.3 The intuition of distinctness and the vagueness constraint

According to the account of the phenomenal concept (at least Papineau's account, which claims that the phenomenal concepts are vague because they have no a priori connections with material properties), the quality space model is incompatible with the phenomenal concept strategy because some versions of the quality space model, e.g. Clark (Clark 2000), presuppose that there could be a direct mapping between phenomenal properties and material properties and that the issue of the uniqueness of the semantics of phenomenal concepts can be bypassed in that way. But can we make a use of a model that is only technically inspired by quality space/the quality space model but which differs significantly inasmuch as it tries to map different psychological concepts, that is, the

concepts of consciousness that are based on the descriptions of the causal/functional roles and itself can be defined in terms of their own quality spaces? Can we use such a model in explaining away the so-called “intuition of distinctness”? Perhaps we can.

Even if Chalmers, Levine and Jackson were right about the universal applicability of the a priori entailment thesis to reductive explanations of natural kinds, it wouldn’t have any bearing on psychophysical identity because it relies on a completely different explanatory model. Namely, we do not explain the a posteriori necessary status of psychophysical identity by an appeal to descriptions of causal roles, as we do in the case of natural kinds, we do it in terms of the roles of proper identities in intelligible explanations. Since phenomenal concepts use direct modes of presentation and therefore cannot have a priori connections with material properties, they are ultimately vague in their nature. This, however, requires them to have an account which explains why one would assume, in the first place, that they are supposed to use descriptions of the causal roles as modes of presentation in order to be fully explanatory. This in my understanding is the “intuition of distinctness”. Perhaps we could explain this intuition away by studying the psychological concepts used in this area and their links to other concepts of consciousness that are also based on some kind of description of causal or functional roles. Perhaps by studying specifically psychological concepts instead of the standardly used natural kind concepts we could come closer to completely removing the sense of puzzlement associated with psychophysical identity based on proper phenomenal concepts. For such a task I propose to use a variant of the quality space model that would not map phenomenal properties directly onto material properties, but instead analyze different psychological concepts within a quality space. The structure of the resemblance and the difference in each of these domains should paint a clearer picture of why phenomenal concepts are also expected to refer only by using descriptions of causal roles as modes of presentation. In this way the intuition of distinctness can be explained away.

The intuition of distinctness is a complex notion. It can be formulated as a specific issue as opposed to a general claim of dualistic intuitions. We could say that it is an intuition according to which all intelligible explanations must rely on the notion of necessity stemming from a priori derivation from some more basic level such as microphysics. In my understanding, it is the intuition according to which reductive explanations or identities

must be based on modes of presentation that involve descriptions of causal roles in order to be fully explanatory. As we remember from the last section of the previous chapter, Levine, Chalmers and Jackson maintain that psychophysical identity must be based on such modes of presentation because, in their opinion, necessary statements require a priori derivation from (complete) microphysical descriptions or a priori analysis of meaning. As it has been shown in the previous chapter, no such analysis is required for the cases that involve the identities of natural kinds so there is no reason that it should be required for psychophysical identity. Furthermore, a requirement for such analyses does not stem from the explanatory gap account or from anti-physicalist arguments alone, it is rather stipulated as a tacit assumption in those arguments and therefore can be considered as nothing more than an intuition.

As it was demonstrated in the third chapter, the explanation of the intuition of distinctness may take different forms that are not mutually exclusive. Papineau (Papineau 2002) offers an explanation of this intuition in terms of a conceptual fallacy and calls it the "antipathetic fallacy", that is a fallacy of failing to recognize that it is coherent to think that a concept can have two different modes of presentation. Papineau's main motivation for favouring conceptual dualism over property dualism, as we have seen, is based on the causal argument (Papineau 2002). I concur, that is a very persuasive reason, but perhaps it is possible to reinforce it with an analysis of the concepts in other areas relevant for the study of consciousness. In that way we get an independent argument for favouring conceptual dualism over other alternatives.

Furthermore, I also believe that we need something like a quality space model that would at least rule out qualia inversion cases. Phenomenal concepts are incapable of doing that simply because of their vague nature, i.e. because they don't refer via descriptions of the causal roles. If one were to consider a scheme of causal roles, one would not be able to find phenomenal concepts in there. Of course, we can stipulate their correlation but that would not prevent undetectable qualia inversion cases.

As we shall see in a moment, the key problem that the standard Clark's model of quality space fails to address is how it is possible that psychophysical identity can be a posteriori necessary. He seems to be merely presupposing that physicalism is true and

basically develops a model of psychological concepts. He distinguishes between sensory properties and qualitative properties but also maintains that these should not be connected with particular conceptual content. This distinction places his whole approach within the domain of psychological concepts because on this view direct mapping between sensory properties and qualitative properties is supposed to bypass the issues of causality but still act as a scheme of the functional roles.

The quality space model is meant to address the issues surrounding sensed colour location as well as the problem of metamers, objects with different wavelengths of reflectance that we perceive as the same colour. Obviously, defining colours solely in terms of wavelengths does not cut the mustard here because, as we will ascertain when considering Rosenthal's account of quality space, such a model would be single dimensional and would not allow for properly distinguishing properties on the axis.

Aside of the limitations of the direct empirical study of phenomenal concepts due to their peculiar vague nature, the quality space model can provide valuable information about which theory of consciousness is more likely to be true.

Before trying to answer these questions let me first examine what I consider to be a generic model of quality space.

5.4 Clark's model of quality space

The best illustration for the idea of a quality space can be found in the quality space model for colour vision proposed by Austen Clark (Clark 2000). Roughly put, the quality space model uses the geometrical ordering of sensory properties to explain how phenomenal properties and qualitative properties match and how we come to the structure of their ordering. For example, according to Clark (Clark 2000) the quality space model answers the spectrum inversion arguments by making any such inversion detectable. Although Clark's quality space model is envisaged with, in my opinion, wrong aim, that is, to directly map phenomenal properties onto qualitative properties thus bypassing the issues of the phenomenal concepts' semantics, it might serve as an independent account of the

intuition of distinctness involved within psychophysical identity. On the other hand, although quality space has been mainly developed for the study of colour vision and emotion, we could also apply it to pain by modifying Aydede and Price's experiential-phenomenological approach (Aydede and Price 2006). To make this idea clearer, I need to lay out the original quality space model in some detail.

The quality space model "Clark" style tries to avoid issues of conceptual semantics by directly mapping the phenomenal properties to neurobiological properties in the quality space. On this view then, we don't even need to know what the phenomenal properties are, for, however we define them beforehand, it would not matter because they are identical with the physical or neurobiological properties ontologically. We only need to find their place in the quality space and this is how we explain their a posteriori necessity.

Clark suggests that the nature of the phenomenal could be mapped within the quality space constituted by the multi-dimensional geometrical space that consists of arrays of hue, saturation and brightness. It is very important how we interpret the qualities in the space. The model hinges strongly on the distinction between phenomenal properties and qualitative properties. Following Sellars (Sellars 1963) and terminologically more relevant Strawson (Strawson 1989), Clark (Clark 2000) distinguishes these two notions in the following way: phenomenal properties characterize how things that are objects of experience appear, whereas qualitative properties are the properties of the internal states in virtue of which objects of experience appear as they do. Furthermore, internal states that qualitative properties are properties of, as it were, determine phenomenal properties because things appear to us in a way they do in virtue of these internal states.

Quality space should not be understood as being constituted by the stimuli, but rather by the qualities stimuli present. In a word, same stimulus may present different qualities, or disparate stimuli might present the same quality in quality space. That is why the quality space cannot consist of stimuli. The main point is that the differences among qualities cannot be cashed out in terms of differences in stimuli. It is rather the case that we look for differences within the quality space and their ordering along the geometrical axes that constitute quality space. Clark suggests that the way in which qualities are ordered in the geometrical space is determined by the features of appearance that generate the

structure of ordering. These are matching, discriminability and relative similarity. On this account it would be expected that qualities that are more similar to one another, fall closer in the quality space. Discriminations among stimuli help to order the qualities that stimuli present.

I am confident that the quality space model can be constructed for any kind of experience including pain.

For example, a multi-dimensional quality space of pain might include several geometrical axes of perceived phenomenal qualities, the McGill pain questionnaire (Melzack and Torgerson 1971), sensory axis, behavioural axis and cognitive axis. Furthermore, each of these axes can be described by its own quality space model, for example, the sensory axis might have several axes: spatial, that maps the body location, temporal that marks timing of pain, which in turn can be explained by the type of nerve fibres and myelinization. In this way we get an outline of the structure of experience which in effect is fortified by the perceptual/phenomenal concept semantics and which can help to understand: on the one hand, how the intuition of distinctness stems from the structure of experience and why its grip is so strong; and on the other hand, what makes it intelligible that, metaphysically, there is only one property under two different modes of presentation.

To fully appreciate the idea of quality space let's consider another account of the quality space model (Rosenthal 2010).

5.5 Rosenthal's account of the quality space model

Rosenthal makes a very interesting point (Rosenthal 2010) about the intuitions underlying the explanatory gap and the conceivability arguments. He argues that these problems rely on a very impoverished conception of consciousness, which only takes into account consciously aware access to mental qualities but says nothing about qualities of which we are not consciously aware but which affect our mental processing nevertheless. He finds a support for these claims in cases such as blindsight and masked priming, both of which are based on cases in which people find themselves unable to report perceiving

anything but still react or behave in a way as if they actually perceived something[44]. If we based our conception of what is conceivable solely on how things appear in experience, we are at risk of conflating what merely appears conceivable with what is actually conceivable. This further means that conceivability of qualia inversion cases rests on an impoverished conception of conceivability that only operates with the consciousness based notion of conceivability.

Rosenthal argues that we have access to mental qualities not only through consciousness, but also through their role in perception. According to this distinction he classifies theories of mental qualities into two groups: the consciousness based and the perceptual role based. The consciousness based theories claim that the only way we have access to mental qualities is through consciousness, whereas perceptual role theories claim that mental qualities are the properties in virtue of which we make perceptual discriminations. Perceiving then warrants that the inversion cases of mental qualities cannot occur undetected and it does so because the structure and ordering of mental qualities matches the structure and ordering of perceptible properties, not the individual properties on either side. Provided that perceiving can also occur at the unconscious level it also warrants that this detection is made from the third person perspective, or at least independently from the first person perspective. This is a tremendous advantage of the quality space model because, as we recall, the qualia inversion cases were thought to be possible only because we had no other access than conscious access, i.e. only access from the first person perspective. Empirical evidence that supports the claim that discriminating mental qualities can occur without conscious awareness include cases of blindsight and experiments with masked priming, in masked priming subjects cannot report seeing anything, but nonetheless the stimulus affects their mental processing. (Rosenthal 2010, p. 7), in the case of blindsight, something very similar happens. Due to some injury or damage in the V1 primary visual cortex, people report not being able to perceive any visual stimuli, but are nonetheless able to react to presented objects or to navigate through the room as if they were seeing them. On this understanding, an access to mental qualities can be completely independent from consciousness and in fact can occur in the absence of

[44] I will explain these in a short while below.

conscious awareness of those qualities altogether. This feature of the perceptual role approach to mental qualities has essential links to perception-not just to consciousness.

On the other hand, the consciousness based approach is the familiar approach, we are able to discriminate among stimuli only by way of their appearance through the first person perspective. The perceptual role approach instead takes into account the ordering of the mental qualities in virtue of which we are able to discriminate among perceptible properties, according to their relative similarity. Rosenthal maintains that the perceptual role approach is superior to the consciousness-based approach because it can account for the qualities usually not consciously accessible but still affecting our mental processing and it also prevents undetectable qualia inversion cases.

In my opinion, Rosenthal's approach seems to fall into a general category of psychological concepts, because it only deals with descriptions of functional roles. I also agree with Rosenthal about the claim that intuitions that are used as data in thought experiments and conceivability arguments encapsulate theories as opposed to providing independent support for them. Furthermore, the explanatory gap does not only concern intuitions but, more importantly, how the concepts in this area work. Although in my opinion intuitions play an important role as tacit assumptions in these arguments, the primary target should be explaining the epistemic gap and thus solving the explanatory gap. The epistemic gap, that I have discussed in much greater detail earlier, concerns the conceptual semantics employed when we think about mental qualities, or in terminology accepted in this work: phenomenal properties, from the first person perspective and from the third person perspective. The explanatory gap arises when we cannot explain the epistemic gap. The real challenge then is to explain how physicalism could be true given the epistemic gap, i.e. how conceptual dualism could be true. Although I agree that the phenomenal concept strategy would probably fall within the consciousness-based approach, I think that Rosenthal's general approach deals only with the psychological concepts, and thus cannot appropriately address the explanatory gap problem. As we remember, phenomenal concepts have no a priori connections with material properties, and thus are vague in their very nature. If we try to use a framework according to which mental qualities and perceptible properties are ordered according to their relative similarities and differences in the quality space, we are

basically dealing with descriptions of functional roles, this immediately places the whole approach in the domain of psychological concepts.

On the other hand, I also agree with Rosenthal that we need a theory that will account for both conscious and unconscious perception. However, given that this whole approach is only about the psychological concepts I will not go into details of the other arguments in Rosenthal's paper, because it would lead us astray from the discussion I want to pursue here. I would instead like to focus on his account of the quality space that I find really invaluable for my own arguments. He makes two really important distinctions. The first one concerns the general nature of the quality space model. He puts it like this:

"…Nor must the quality spaces for mental qualities reflect properties of the neural processes that subserve, or may even be identical with, those mental qualities. The relevant quality spaces reflect only the similarities and differences among mental qualities determined by a creature's ability to discriminate among various perceptible properties"(Rosenthal 2010 p 11).

In this way he assures that the quality space model is not about the direct mapping of individual phenomenal properties onto individual physical properties[45], but rather an independent method for their detection. As I said, this approach would fall into the category of psychological concepts, regardless of whether we talk about conscious or unconscious perception[46]. It tries to avoid taking into account the phenomenal concepts, which I think is a good strategy given their vague nature and the implausibility of direct mapping, and instead focuses on an independent account of the description of the functional roles phenomenal and physical properties play in a quality space, which in my opinion is the right way to deal with psychological concepts. And this is of invaluable importance for my own general approach.

His other point concerns the intuitions underlying the explanatory gap and the conceivability arguments. I could not put it better than Rosenthal himself:

---

[45] Which as I argued cannot be done due to the vague nature of phenomenal concepts.

[46] Because it only deals with the descriptions of the functional roles in either way.

"Perhaps the most significant contribution of experimental philosophy, then, is to underscore the likelihood that various intuitions are not theory neutral, but products of such theoretical commitments. If intuitions aren't widely shared by the folk and fit conveniently with theories espoused by those who invoke the intuitions, it's natural to ask whether the intuitions encapsulate those theories, rather than providing independent support" (Rosenthal 2010, p. 5).

This passage brilliantly encapsulates the whole idea of my project. Although I don't start with some method or already available research from experimental philosophy that would undermine the explanatory gap intuitions, I instead start by analyzing the anti-physicalist arguments and showing that there is nothing in the arguments themselves that prevents us from drawing conclusions that contradict each other[47]. We can leave it for some other occasion to discuss his other claims in the paper and focus now on his account of the quality space model.

Rosenthal's basic idea of the quality space model can be expressed as the claim that it is a function of discrimination between two perceptible properties. To discern between perceptible properties a creature must be in some states whose differences in virtue of which a creature discerns perceptible properties reflect the differences between perceptible properties. It is important to note that for Rosenthal (Rosenthal 2010) these states need not be conscious, as there is a tremendous amount of empirical evidence about cases in which people are able to make discriminations of which they are not consciously aware. The quality space model captures the differences and similarities between perceptible properties and discriminatory states and gives us a structure of relations within these two domains that can be mapped onto one another.

The point at which Rosenthal departs from Clark's standard model is that he includes spatial properties such as size, shape, and distance as well. Now, the spatial properties perceived by different sensory modalities, such as shape, colour, distance and location, are the discriminations of the same perceptible property, however, the states we go into when discerning them are different.

---

[47] See the chapter four.

The quality space model according to Rosenthal's interpretation is not meant to show how mental qualities or the phenomenal character/properties map onto the perceptible properties, but rather how the model reflects a creature's abilities to make qualitative discriminations, consciously aware or not. On this interpretation, the quality space model is supposed to accommodate the resemblance between mental qualities and perceptible properties that is based on a creature's ability to discriminate qualities in the first place. Furthermore, the model then is not committed to denying the existence of two types of properties but rather to accommodating the structure of resemblance between the two domains. The main point of this interpretation of the model is that there is no resemblance between individual properties but rather a resemblance at the level of families of properties (Rosenthal 2010, p 12). The central point in the Rosenthal's interpretation of the quality space model is the distinction between a consciousness role and a perceptual role of qualitative discriminations. Rosenthal rightly argues that the quality space model renders qualia inversion cases nonsensical because, given the asymmetrical topology of quality space, undetectable inversions are impossible.

As we remember, qualitative discriminations occur on both consciously aware and consciously unaware levels. The model maps the structure of resemblance between properties on either side, but not individual properties onto each other. This feature of the model precludes undetectable qualia inversion cases. The key point here is that all discrimination according to Rosenthal can occur both at the unconscious level and at the conscious level. However, the intuitions about conceivability of zombies and about the explanatory gap pertain only to the conscious level, so the quality space model must account for the discriminations occurring at the conscious level.

On Rosenthal's view (Rosenthal 2010) the quality space does not rely on the discrimination of the physical properties of the stimuli, e.g. wavelengths of light or frequency of sound, etc, for such a quality space would be single-dimensional and the properties of either side of the axis would seems indistinguishable, but rather reflects relative similarities and differences among mental qualities that the creature is able to discriminate among perceptible properties (Rosenthal 2010, p. 11).

If there were nothing about the mental qualities but their location in the quality space, then visual and auditory space would seem identical/indistinguishable (Rosenthal 2010, p. 13).

In the final section of his paper Rosenthal considers higher order awareness (HOA) to argue that the perceptual role theory in combination with the HOA is more plausible than the consciousness based one, because it accounts for the perceived but consciously not registered stimuli. But as I said I will not be disusing that part of his paper because it is out of scope of my argument.

Having presented a general framework in which the empirical study of phenomenal coconsciousness can proceed I turn now to briefly examining some details of how the quality space might look for pain. I believe that a similar approach can be used for other touchstones of phenomenal consciousness as well.

5.6 The quality space for pain

The quality space model encounters problems straight off when one tries to apply it to other touchstones of phenomena consciousness. To begin with, the model should be universally applicable to all experiences that involve something it is like to be in that state/having that experience. Clearly, there is something it is like to be in pain or undergoing certain emotion. Is this model universally applicable to these two kinds of states as well? David Rosenthal points out[48] a quality space of emotion would need to involve intentional content in addition to qualitative character, which would clearly make it more difficult than the standard cases of visual perception, but feasible nevertheless. What about pain? Although there is something it is like to be in pain, there is some quality associated with pain, it is not clear at all whether pain is a kind of perception and whether mapping of qualitative properties onto perceptible properties even makes sense in this case. It should not be controversial that if it were possible to construct a quality space model for pain, then

[48] He remarked this in a correspondence.

it would be capable of accounting for pain dissociation syndromes like painfulness without pain and pain without painfulness (Grahek 2007), as it is capable of detecting spectrum inversions in the classical Clark's model (Clark 2000) and in Rosenthal's model (Rosenthal 2010). In this case the detection is only at the conscious level, but, as in the inverted spectrum cases, it should involve an integration of the first person and the third person approaches. There are several issues associated with constructing a quality space model for pain. Here is the list of some the issues:

a) What are the properties in virtue of which we are able to make discriminations among perceptible properties of pain?

b) What are the perceptible properties of pain?

c) How many axes the quality space of pain there should be and what they should consist of?

d) What are the modalities of pain?

e) Can we construct a quality space for all modalities of pain?

I won't try giving a definitive and comprehensive answer to these questions because that would require another whole book. I would rather like to point out some issues associated with constructing a quality space for pain and only gesture towards what I believe to be the right answer to them.

When it comes to qualitative properties of pain there are several issues associated with it. To begin with, can we have unconscious pain-the pain we are not consciously aware of, like in the case of masked priming or the blindsight? In a word, can we make the same distinction between consciousness based and perceptual role based theories of pain as Rosenthal suggested for colour vision? Does this distinction break down in the case of pain and would it matter if it does? Perhaps this distinction does not matter in the case of pain. As it turns out, pain is an extremely complex phenomenon and studying it in terms of psychological concepts can afford us some really valuable insights about the relation between phenomenal concepts and psychological concepts of pain. After all, the quality space is all about the psychological concepts so even though dissociations between sensory

and affective components of pain happen only at the conscious level, we can still use the quality space to describe the functional roles that properties play.

On the other hand, there are several other aspects of pain that would require their own quality spaces and then each of these quality spaces can enter the generic quality space for pain as a single axis. For example, according to Melzack and Wall (Melzack and Wall 1996, p. 137) there are three main psychological aspects of pain: sensory-discriminative, motivational-affective, and cognitive-evaluative. Essential element in all of these interactions is descending inhibitory control mechanisms, because it influences the overall experience of pain by modulating afferent impulses. Now, each one of these aspects could require its own quality space and also be an axis in a more general quality space.

Speaking of the variations of pain perception, there have been several studies that suggest that the pain perception depends on a cultural background (Bates 1987), and even on previous painful experiences (Melzack and Wall 1996).

Another important issue stems from considerations concerning pain dissociation syndromes. Grahek (Grahek 2007) argues that pain without painfulness and painfulness without pain are the only genuine pain dissociation syndromes. Given that perhaps there is no unconscious pain, after all it is its biological and evolutionary role to alert us and keep us from further harm, what could the axes of qualitative properties of pain consist of? Perhaps we could combine two approaches, that are actually very similar, to determine that. On one approach we determine the multiple clusters of qualities across different modalities of pain by using the McGill pain questionnaire. On the other we use a method of horizontal and vertical techniques of experimentally studying phenomenology to tease out a subjective feel by using Aydede's method[49]. We combine these two axes, which on the other hand could consist of several other axes. The main difference to the quality space for colour vision is that the quality space for pain takes as perceptible properties kinds of tissue and damage instead of physical processes that are responsible for inflicting the damage and are in effect external to the body. The dependence on the kind of tissue and the location of damage in the tissue are highly variable even if kinds of fibres innervating the tissue remain constant.

[49] I will come to this in a short while below.

For example, deep tissue pains are perceived with inaccurate localization and superficial pains are perceived more precisely (Melzack and Wall 1996). To capture these sophisticated difference we need to go into some details of the complexity of studying pain empirically.

Consider for example the gate theory of pain. According to Mezack and Wall (Mezack and Wall 1965) a pain sensation is modulated in the spinal cord before it is processed in the brain and before any motor response can take place. Specifically, they have proposed that the stimulation from the skin evokes neural impulses that go into three physiologically very different areas of the spinal cord. The impulses are sent to: cells of the substantia gelatinosa in the dorsal horn, the dorsal –column fibers that project toward the brain and the first central transmission cells (T) in the dorsal horn (Melzack and Wall 1965, p. 974). Melzack and Wall propose that the substantia gelatinosa serves as a gate control system that modulates ascending nerve impulses before they reach and affect the T cells. The ascending patterns in the dorsal column control triggering and activation of the selective brain processes that modulate or influence the gate control system. Finally T cells activate neural mechanisms that are responsible for response and perception (Melzack and Wall 1965, p. 974). The perception of pain evoked by skin stimulation, in their view, is determined by these three systems. In this way the gate controls both the bandwidth of the gate and also impulses coming from the brain as a response to noxious stimuli. This further means that there is an inhibitory and excitatory system of pain in place that controls how we perceive the pain stimuli. So what are the properties in virtue of which we are able to make discriminations among perceptual properties of pain? It is a difficult issue, but perhaps we could get closer to answering it by examining some proposals on how to use introspective reports about quality of pain in clinical settings.

Consider for example the Aydede and Price method of experimental procedure that scientists can use themselves when engaging into the introspective experiments (Aydede and Price 2006). Instead of experimental they are calling it the *experiential* approach, because essentially it relies on the first person experiential reports. They divide their experimental method into two phases: the horizontal phase and the vertical phase. In the horizontal phase introspective reports of the trained subjects are aimed at characterizing relations of the overall conscious experience without attempting to relate them to brain structures (Aydede and Price 2006, p. 259). In the vertical phase experiments try to discover

relations between the brain activity and the results from the first phase. For example, subjects are asked to undergo the same stimulation procedure while scanned at the same time. This approach is destined to failure if Aydede and Price believe that it can directly bypass the conceptual issues of the explanatory gap and especially skip over the vagueness constraint of phenomenal concepts. They seem to be arguing exactly for that. However, I find their approach useful because it can reinforce the quality space for pain as long as it is understood that it pertains only to psychological concepts.

Aydede and Price's horizontal/vertical approach can suitably accompany clinical and diagnostic methods such as the McGill pain questionnaire (Melzack and Torgerson 1971; Melzack 2005; Dubuisson and Melzack 1975), which is basically used in clinical settings for measuring the quality of pain. Melzack and Torgerson have proposed to categorize and scale common words used to describe pain according to their intensity and dimensions. What they have found is that categories and dimensions in different people are grouped in the same relative dimensions. Words that describe pain are categorized into several classes according to their modality. For example, one class consists of words describing *sensory* quality of pain (e.g. descriptions in terms of spatial, thermal, temporal, pressure and similar properties), the other class is about *affective* qualities of pain, e.g. tension fear and such; the third class is the *evaluative* one, it consists of words describing subjective overall intensity of pain experience (Melzack and Torgerson 1971, p. 51). They have run an experiment with diagnosed patients and the result was that words from different dimensions tend to consistently cluster together in a proto quality space for pain[50]. This model only seems to capture the qualitative and subjective side, although one of the dimensions is about sensory discriminations, it is only about their perceived quality. The model does not mention what Rosenthal calls "perceptible properties". Perhaps the three dimensions of this quality space, i.e. sensory, affective and evaluative can be accompanied by the phenomenal-phyiscal dimension proposed by Aydede and price (Aydede and Price 2006) and comprise an axis in a multi-dimensional and multi-level general quality space for pain that would allows us to zoom into and out of the actual descriptions of the functional roles for different modalities and dimensions of pain thus further allowing us to see why phenomenal concepts are

[50] They don't call it quality space, but basically it has a lot of the elements of a quality space.

expected to refer via descriptions of causal roles. This would explain away the intuition of distinctness.

Another important aspect of pain is its sensory/physical location and intensity. That is to say, issues of the perceptible properties of pain. Perhaps we could think of it in terms of specialized and specific neural fibers instead of in terms of damage or in terms of noxious stimuli. There are three major groups of fibers: A-beta or large myelinated, A-delta or small myelinated, and C or unmyelinated fibres. Each one of these transmits impulses at different speeds[51] which affects the sensation of pain. The intensity of pain can be thought of in terms of on the one hand specific and on the other specialized nerve fibers. The difference between the two comes down to this: specific fibers usually perform a single or unique task. On the other hand specialized fibers can perform several kinds of related tasks. However, some studies using neuronography[52] have shown that there is no simple relationship between a type of fibre and a quality of sensation (Melzack and Wall 1996, p. 157). That is to say that there are no "pain" fibres, but only fibres specialized to respond only to the intensity of stimulation. This further means that the quality space for pain might also come very useful in studying this kind of issue.

The list of issues grows exponentially when one is faced with such a complex phenomenon as consciousness. However, the case of quality space for pain shows that psychological concepts in this area, that are based on descriptions of functional roles, do not necessarily work in the same way as physical concepts or concepts of natural kinds. Considerations about the quality space and psychological concepts bring us closer to an understanding of why phenomenal concepts are expected to refer via descriptions of the causal roles and, more importantly, why they should not be expected to refer in that way.

Considered in this way it becomes apparent that the whole point of studying psychological concepts instead of natural kinds and physical concepts is that psychological

---

[51] Depending on the myelnization. Myelinated fibres transmit impulses faster than the unmyelinated ones.

[52] Neuronography is an experimental technique of tracking fibre activity and perceived sensation by inserting microelectrodes into human nerves and recording the impulses in all the sensory afferent pathways.

concepts can reveal unique complexity of consciousness phenomena while also preserving the vagueness of phenomenal concepts.

Conclusion

I began my discussion with the considerations about what is the proper way to even start thinking about consciousness. I argued that instead of discussing the available doctrines beforehand, we should ask ourselves what are the most basic assumptions that we can make about consciousness. Once we formulate these assumptions we will be able to come up with the most plausible characterization of the way to think about consciousness. I have discussed Papineau's *causal argument* (Papineau 2002) and concluded that the most plausible doctrine about consciousness is physicalism thought of as an identity thesis according to which mental states are identical with brain states or some higher order physically realized states. Once I settled the starting position I continued by examining some of the most important arguments against this position. In the first two chapters I restrained myself from providing any extensive criticism of these arguments because I wanted to make it as clear as possible what are the original claims. I have decided to approach to these issues by gradually branching out towards a "big picture" and develop my own position steadily. Of course I occasionally give out my own position in the first two chapters, but I offer really thorough criticism from the third chapter onwards.

In the first chapter I begin with Nagel's account of the epistemic gap (Nagel 1974). He claims that the main feature of consciousness is that there is something it is like to be conscious for a given creature. This "what it is like" aspect of consciousness will always be left out from any functionalist or reductive explanation of consciousness because it can only be known from the first person perspective of a given sentient being. That is to say, a theory of the subjective aspect of consciousness is always tied to a certain point of view, namely the first person perspective, which cannot be eliminated from a reductive explanation in the same way as it can in the case other natural phenomena. Here we have a version of the epistemic gap that I already mentioned. He basically claims that there are two epistemically distinct ways of knowing something. Knowing things in one way does not enable one to know things the other way. That is, we can know about the neural underpinnings of conscious experience from the third and objective point of view, but to conceptualize it from the first person perspective we need to be that very organism.

The proto-knowledge argument in Nagel's account can best be seen in the claim that the knowledge of all the physical or functional facts about the brain does not entail facts about the what-it-is-like of experience, for such a knowledge would require one to become a creature whose experiential perspective we are talking about. As I said earlier, I didn't discuss or criticize in greater detail the arguments in the first chapter because I wanted to present them in their original form. I gave out hints here and there about my own position in these matters, but I don't develop them until chapter three and after it.

After that I go on and present the Knowledge Argument (Jackson 1982, 1986). The argument formally looks like this:

"13. Mary (before her release) knows everything physical there is to know about other people;

14. Mary (before her release) doesn't know everything there is to know about other people (because she *learns* something about them on her release);

Therefore,

15. There are truths about other people (and herself) which escape the physicalist story" (Jackson 1986, p. 293).

From this Jackson concludes that the fact she learns upon release cannot be based on some physical properties, because if there were physical, she should have known them while still in the room. Thus if phenomenal properties exist and are not physical, then epiphenomenalism must be true concludes Jackson (Jackson 1982).

I argued in chapters three and four that this conclusion is not warranted by the Knowledge Argument itself, and it fails to establish that the physical and the phenomenal are metaphysically distinct. To do that, he needs some further tacit assumptions about the way in which concepts operate. He seems to be relaying solely on the intuitions in this case, which makes the argument very problematic.

In the third section of the first chapter I presented Kripke's modal argument against the identity thesis (Kripke 1980). His basic point is that identities must be, if true at all, necessary

true. Psychophysical identity is supposed to be a posteriori necessary like any other identity of the natural kinds. However, in the case of natural kinds identities we can conceive of their falling apart, for example that heat is not identical with molecular motion. If we can conceive of their falling apart they are not necessary but contingent on Kripke's account. He says that this merely seems contingent and we explain it away as a misdescription. What we actually mean by this is not that molecular motion is not identical with heat, but that something other than molecular motion could have produced a sensation of heat. However, such an analysis is not available for psychophysical identity, because what seems like pain is in fact pain, so when we conceive of their falling apart they really fall apart, we cannot dismiss it as a misdescription. Here we have a very important point about inference from conceivability to metaphysical possibility. This point has been at the heart of many vigorous debates in philosophy of mind and metaphysics and it has been a common ground for attacking Kripke's argument. Furthermore, this issue is of central importance for the explanatory gap account, because as we saw in the second chapter, the epistemic gap does warrant only entailment from conceivability to epistemic possibility.

In the next section I discussed further attempts to fortify this position by David Chalmers (Chalmers 1996). He tried to justify Kripke's conclusion by introducing the argument based on the supervenience thesis and two-dimensional semantics. Chalmers's account can be summed up in this way: if the phenomenal is not logically supervenient on the physical, and the physical domain is causally closed, then phenomenal properties are distinct properties from the physical properties. He provides another independent argument for the claim that the phenomenal is not supervenient on the physical, i.e. the zombie argument. However, it does not follow from his arguments alone that supervenience based characterization of physicalism is better than identity based characterization. I won't be discussing his account of the two-dimensional modal semantics here, because I already did that on two different occasions in the dissertation[53].

In the second chapter I discuss the explanatory gap account (Levine 1983, 1993, 2001). I tried to follow Levine in all sophisticated distinctions he makes and I discussed notions of

[53] See the last two sections of the first chapter and the last section of the fourth chapter.

thick and thin concepts and thick and thin conceivability, gappy and non-gappy identities, and so forth. His distinction between thick and thin concepts is very important for the phenomenal concept strategy because it appeals to the crucial difference between phenomenal concepts and the natural kinds concepts. That is to say, thick concepts are like phenomenal concepts, substantive and determinate about their own content, although phenomenal concepts are vague considered from the third person perspective because they don't have a priori connections with the material properties. Thin concepts are like natural kinds concepts, they are not determinate and substantive. Levine argues that identities containing thick concepts cannot be fully explanatory, as they cannot account for the twofold task of intelligible explanations. That is to say, gappy identities, cannot identify explanans with the explanandum because thick concepts use direct and determinate modes of presentation and thus cannot appeal to the descriptions of the causal or functional roles in identifying the explanans and the explanandum. I argued at the end of this chapter that an explanation of the psychophysical identity containing thick concepts (phenomenal concepts) only seems unintelligible because of the tacit additional assumptions. After all, if Levine accepts that physicalism formulated as an identity thesis is true, then he also have to accept that identities are not explanatory.

In the third chapter I finally come to the point when I could discuss the main and most plausible response to the explanatory gap-it is the phenomenal concept strategy (Stoljar 2005). According to the strategy phenomenal concepts are a special kind of concepts, which unlike natural kinds concepts refer directly, or as Hill and McLaughlin put it:

"The reference-fixing property that is associated with a theoretical concept is identical with the property to which the concept refers. " (Hill and McLaughlin 1999, p. 452)

I then go on and discuss several different accounts of phenomenal concepts. I begin with Loar's account of recognitional phenomenal concepts (Loar 2002) and finish with Papineau's account of phenomenal concepts as a subclass of perceptual concepts that only use a linguistic form of a demonstrative but are not themselves demonstratives (Papienau 2007). In this chapter I also discussed the inconsistency objection to the Knowledge Argument, i.e. that one cannot coherently hold both epiphenomenalism true and phenomenal properties causally inefficacious. If they were in fact causally inefficacious we

would never be in a position to detect them or to know about them. This argument is supposed to corroborate the first premise of the causal argument (Papineau 2002). In the rest of the third chapter I discuss two important objections to the phenomenal concepts, one is "Max Black's objection" (Block 2007) and the other is physical explicability and explanatory adequacy of phenomenal concepts. Max Black's objection consists of the claim that two different modes of presentation must reflect two distinct properties. I argued along the same line as Ned Block (Block 2007) that such objection presupposes that the difference in cognitive modes of presentation entails the difference in metaphysical modes of presentation, which does not have to be the case. I consider an example of a student of French language who learns that a French word for "cat" is "chat" by being shown the same cat as an illustration on two different occasions and thus he thinks there are two different meanings of "chat". The example shows that actually there is a difference in cognitive modes of presentation associated with each occurrence of the same cat, but the difference stems from the one and the same metaphysical mode of presentation. The second objection concerns physical explicability and explanatory adequacy of phenomenal concepts. Chalmers (Chalmers 2007) argues that if phenomenal concepts were physically explicable then they are not able to explain the explanatory gap and vice versa. I argued that Chalmers's argument does not undermine phenomenal concept strategy because it merely restates the explanatory gap problem. Namely, phenomenal concepts refer directly and therefore cannot have a priori connections with the material properties. If they had a priori connections with the material properties there would be no epistemic gap and consequently no explanatory gap.

In the fourth chapter I developed my key argument. I argued that although there is the epistemic gap it is not inexplicable. Some philosophers (Chalmers 1996, 2007; Chalmers and Jackson 2001; Levine 1983, 1993, 2001) argue that an explanation of the epistemic gap must come from the identity itself. I however argue that it doesn't have to be the case, especially if we take into account vague nature of phenomenal concepts, or as I call it in the fifth chapter the vagueness constraint. In this chapter I also argued that the conceivability arguments and the explanatory gap account rely on some intuitions that are in fact in need of explanation. The intuitions in question are about the requirement that phenomenal concepts have to use descriptions of the causal or functional roles as modes of presentation in order to be fully explanatory. I called this issue a brute disagreement about intuitions and

argued that it cannot be resolved by the available arguments but by an account that would explain those intuitions away. I urged that the issue at stake is a version of Papineau's notion of "intuition of distinctness" (Papineau 2002). At the end of this chapter I discussed the role of a priori derivation and a priori knowledge in the issue of intelligibility of explanation.

Finally, in the fifth chapter I examined a prospect for the empirical study of phenomenal consciousness[54]. I started off with discussing the vagueness constraint of the phenomena concepts for the empirical study of phenomenal consciousness. As I have concluded in the fourth chapter, phenomenal concepts have no a priori links to the material properties and therefore are vague. Another way to put this would be to say that phenomenal concepts use direct modes of presentation and thus cannot be directly mapped in the scheme of the causal or functional roles. Instead of trying to find directly the corresponding material properties of the phenomenal concepts I argued we should use a quality space model to analyze psychological concepts in this area and try to understand why are phenomenal concepts expected to refer via descriptions of the causal or the functional roles. I discussed Clark's version of the quality space for colour (Clark 2000) and Rosenthal's account of the quality space (Rosenthal 2010) and then proposed my own account of the quality space for pain. I argued that the quality space for pain would be a real challenge but nevertheless feasible. Some of the main problems with the quality space for pain would be discerning how many axes it should have and which modalities should be represented in it.

Clearly, this is not the last word on this topic, but the reader should take two very important morals from my discussions. The first one is that we cannot use empirical evidence to solve the philosophical problems directly. And the second is that the most reasonable way to use empirical evidence when considering philosophical problems is to analyze how the specific explanations work in a relevant area and how specific concepts in these areas relate to other relevant concepts.

---

[54] I insist here on the specific aspect of consciousness, and that is why I write "phenomenal consciousness" instead of just "consciousness" because empirical study of some other aspects of consciousness is not that intractable as in the case of phenomenal consciousness. I already mentioned Chalmer's distinction between hard and easy problems of consciousness. That is why I think it is important to emphasize that we are talking about the phenomenal consciousness.

References:

Addis, L., 1986. Pains and Other Secondary Mental Entities. Philosophy and Phenomenological Research, 47*(1),* pp. 59-74.

Akins, K., 1993a. What is it like to be boring and myopic? In: B. Dahlbom, ed. 1993. Dennett and his Critics. Oxford: Blackwell.

Akins, K., 1993b. A bat without qualities? In: M. Davies and G. Humphreys, eds. 1993. Consciousness: Psychological and Philosophical Essays. Oxford: Blackwell.

Alter, T., 2002. Nagel on imagination and physicalism. Journal of Philosophical Research, 27, pp. 143-158.

Alter, T. and Walter, S. eds., 2007. Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism. Oxford: Oxford University Press.

Armstrong, D. M., 1962. Bodily Sensations. London: Routledge and Kegan Paul.

Armstrong, D. M.,1964. Vesey on Bodily Sensations. Australasian Journal of Philosophy, 42, pp. 359-362.

Armstrong, D. M., 1968. A Materialist Theory of the Mind. New York: Humanities Press.

Aydede, M., 2000. An Analysis of Pleasure vis-à-vis Pain. Philosophy and Phenomenological Research, 61*(3),* pp. 537-570.

Aydede, M. and Güzeldere G., 2002. Some Foundational Issues in the Scientific Study of pain. Philosophy of Science, 69*(Suppl.),* pp. 265-283.

Aydede, M. and Güzeldere G., 2005. Cognitive Architecture, Concepts, and Introspection: An Information-Theoretic Solution to the Problem of Phenomenal Consciousness. Noûs, 39*(2),* pp. 197-255.

Aydede, M. ed., 2006. Pain: New Essays on Its Nature and the Methodology of Its Study. A Bradford Book. Cambridge Mass.: MIT Press.

Aydede, M.,2006a. A Critical and Quasi-Historical Essay on Theories of Pain. In: Aydede, M. ed. 2006. Pain: New Essays on Its Nature and the Methodology of Its Study. Cambridge, Mass.: MIT Press.

Aydede, M. and Price, D., 2006. The experimental use of introspection in the scientific study of pain and its integration with third-person methodologies: The experiential-phenomenological approach. In: Aydede, M. ed. 2006. Pain: New Essays on Its Nature and the Methodology of Its Study. Cambridge, Mass.: MIT Press.

Baker, L. R.,1998. The first-person perspective: A test for naturalism. American Philosophical Quarterly, 35 *(4),* pp. 327-348.

Balog, K., 1999. Conceivability, Possibility, and the Mind-Body Problem. The Philosophical Review, 108 *(4),* pp. 497-528.

Barber, T. X.,1959. Toward a theory of pain: Relief of chronic pain by prefrontal leucotomy, opiates, placebos, and hypnosis. Psychological Bulletin, 56 *(6),* pp. 430-460.

Barber, T.,1964. The effects of "hypnosis" on pain: A critical review of experimental and clinical findings. Psychosomatic Medicine, 25, pp. 303–33.

Baxter, D. W. and Olszewski J.,1960. Congenital universal insensitivity to pain, Brain, 83 *(3)*, pp. 381-393.

Beckermann, A., 2000. The perennial problem of the reductive explainability of phenomenal consciousness: C. D. Broad on the explanatory gap. In: Metzinger, T.ed., 2000. Neural Correlates of Consciousness. Cambridge, Mass.: MIT Press.

Benardete, J. A.,1958. The analytic a posteriori and the foundations of metaphysics. Journal of Philosophy, 55 *(12),* pp. 503-514.

Berridge, K. C.,1996. Food reward: Brain substrates of wanting and liking. Neuroscience & Biobehavioral Reviews, 20*(1),* pp. 1-25.

Berridge, K. C. and Robinson T. E., 1998. What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? Brain Research Reviews, 28, pp. 309–369.

Berthier, M. L., Starkstein S. E. and Leiguarda R., 1988. Pain Asymbolia: A Sensory-Limbic Disconnection Syndrome. Annals of Neurology, 24 *(1),* pp. 41-49.

Berthier, M. L., Starkstein S. E., Nogues M. A. and Robinson R. G., 1990. Bilateral sensory seizures in a patient with pain asymbolia. Annals of Neurology, 27*(1),* pp. 109.

Biro, J. I. 1991. Consciousness and subjectivity. Philosophical Issues,1, pp. 113-133.

Biro, J. I. 1993. Consciousness and objectivity. In: Davies M. and Humphreys G., eds., 1993. Consciousness: Psychological and Philosophical Essays. Oxford: Blackwell.

Block, N.,1981. Psychologism and Behaviorism. Philosophical Review*,* 90 *(1),* pp. 5-43.

Block N.,1990. Inverted Earth. Philosophical Perspectives*,* 4, pp. 53-79.

Block, N., Flanagan, O. and Güzeldere, G. eds., 1997. The Nature of Consciousness. Cambridge, Mass.: MIT Press.

Block, N.,1996. Mental Paint and Mental Latex, Philosophical Issues, 7*, (Perception),* pp. 19-49.

Block, N. and Stalnaker, R.,1999. Conceptual Analysis, Dualism, and the

Explanatory Gap. Philosophical Review*,* 108*, (1),* pp. 1-46.

Block, N.,2001. Paradox and cross purposes in recent work on consciousness. Cognition*,* 79 *(1-2),* pp. 197-219.

Block, N.,2006. Bodily Sensations as an Obstacle for Representationism. In: Aydede, M. ed. 2006. Pain: New Essays on Its Nature and the Methodology of Its Study. Cambridge, Mass.: MIT Press.

Block, N.,2007. Max Black's Objection to Mind-Body Identity. In: Alter, T. and Walter, S. eds., 2007. Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism. Oxford: Oxford University Press.

Block, N., 2007a. Consciousness, Function, and Representation: Collected Papers, Volume I, A Bradford Book. Cambridge, Mass.: MIT Press.

Blumenfeld, J. B., 1975. Kripke's refutation of materialism. Australasian Journal of Philosophy, 53 *(2),* pp. 151-156.

Braddon-Mitchell, D. and Jackson, F.,1996. Philosophy of Mind and Cognition. Oxford: Blackwell.

Braun, D.,2006. Names and Natural Kind Terms. In: Lepore E. and Smith B. eds., 2006. The Oxford Handbook of Philosophy of Language. Oxford: Oxford University Press.

Bouckoms, A. J.,1994. Limbic Surgery for Pain. In: Wall, P. D. and Melzack, R. eds., 1994. Textbook of Pain. Edinburgh: Churchill Livingston.

Brand, P. W. and Yancey P.,1993. Pain: The Gift Nobody Wants. New York: HarperCollins Publishers.

Broad, C. D.,1959. Scientific Thought. Paterson, NJ: Littlefield Adams.

Byrne, A.,1999. Cosmic Hermeneutics. Nous, 33, pp. 347-383.

Byrne, A.,2001. Intentionalism Defended. Philosophical Review, 110*(2),* pp. 199-240.

Carnap, R. 1947. Meaning and necessity: A study in semantics and modal logic. Chicago Illinois: Chicago University Press.

Carruthers, P., 2001. Consciousness: Explaining the phenomena. Royal Institute of Philosophy Supplement, 49, pp. 61-85.

Carruthers, P.,2004. Reductive Explanation and the 'Explanatory Gap'. Canadian Journal of Philosophy, 34, pp. 153-173.

Carruthers, P. and Veillet, B.,2007. The Phenomenal Concept Strategy. Journal of Consciousness Studies*, 14n (9-10),* pp. 212-236.

Chalmers, D.,1996. The Conscious Mind*.* New York: Oxford University Press.

Chalmers, D. and Jackson, J., 2001. Conceptual Analysis and Reductive Explanation. Philosophical Review*,* 110, pp. 315-360.

Chalmers, D., 2002a. Does Conceivability entail Possibility. In: Gendler T. S. and Hawthorne, J.,eds. Conceivability and Possibility. Oxford: Oxford University Press.

Chalmers, D.,2002b. On Sense and Intension. Nous*,* 36, pp. 135-182.

Chalmers, D.,ed., 2002c. Philosophy of Mind: Classical and Contemporary

Readings. New York: Oxford University Press.

Chalmers, D.,2003. Consciousness and its Place. In: Stich, S. and Warfield, F. eds., 2003. Blackwell Guide to the Philosophy of Mind*.* Oxford: Blackwell.

Chalmers, D.,2004. Epistemic Two-Dimensional Semantics. Philosophical Studies*,* 118 *(1-2),* pp. 153-226.

Chalmers, D.,2007. Phenomenal Concepts and the Explanatory Gap. In: Alter, T. and Walter, S. eds., 2007. Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*.* Oxford: Oxford University Press.

Chalmers, D.,2009. The Two-Dimensional Argument Against Materialism. In: McLaughlin B. P. and Walter S., eds., 2009. Oxford Handbook to the Philosophy of Mind. Oxford: Oxford University Press.

Chrisley, R.,2001. A view from anywhere: Prospects for an objective understanding of consciousness. In: Pylkkanen P. and Vaden T. eds., 2001. Dimensions of Conscious Experience. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Churchland, P. M.,1985. Reduction, Qualia, and the Direct Introspection of Brain States. The Journal of Philosophy,82*(1),* pp. 8-28.

Clark, A.,1993. Sensory qualities. Oxford: Oxford University Press.

Clark, A.,2000. A theory of sentience. Oxford: Oxford University Press.

Clark, A.,2006. Painfulness is Not a Quale. In: Aydede M., ed., 2006. Pain: New Essays on Its Nature and the Methodology of Its Study. Cambridge, Mass.: MIT Press.

Conee, E.,1984. A Defense of Pain. Philosophical Studies, 46, pp. 239-248.

Davidson, D.,1970. Mental Events. In: Foster L. and Swanson J. W. ,eds., 1970. Experience and Theory. Amherst: University of Massachusetts Press.

Davies, M. and Humberstone, I. L.,1980. Two Notions of Necessity. Philosophical Studies, 38, pp. 1-30.

Dempsey, L.,2004. Conscious experience, reduction and identity: Many gaps, one solution. Philosophical Psychology, 17 *(2),* 225-246.

Dennett, D. C.,1978. Why You Can't Make a Computer that Feels Pain, Brainstorms. Cambridge, Massachusetts: MIT Press.

Dennett, D. C.,1988. Quining Qualia. Consciousness in Contemporary Science. In: Marcel A. and Bisiach E., eds., 1988. Consciousness in Modern Science. Oxford: Oxford University Press.

Descartes, R., 1996. Meditations on first philosophy with selections from the objections and replies. In Cottingham, J., ed., 1996. Cambridge texts in the history of philosophy. Cambridge: Cambridge University Press.

Deshpande, D. Y., 1982. An alleged case of factual a priori. Indian Philosophical Quarterly, 9, pp. 107-12.

Devinsky, O., M. Morrell J. and Vogt B. A.,1995. Contributions of anterior cingulate cortex to behaviour. Brain, 118*(1),* pp. 279-306.

Dong, W. K., Chudler E. H., Sugiyama K. and Roberts V. J.,1994. Somatosensory, multisensory, and task-related neurons in cortical area 7b (PF) of unanesthetized monkeys. Journal of Neurophysiology, 72*(2),* pp. 542-564.

Douglas, G.,1998. Why Pains Are Not Mental Objects. Philosophical Studies, 91*(2),* pp. 127-148.

Dretske, F.,1988. Explaining Behavior: Reasons in a World of Causes. Cambridge: MIT Press.

Dretske, F.,1995. Naturalizing the Mind. Cambridge, Massachusetts: MIT Press.

Dretske, F.,1999. The Mind's Awareness of Itself. Philosophical Studies, 95*(1-2)*, pp. 103-124.

Dretske, F.,2003. How Do You Know You Are Not A Zombie? In: Gertler B.,ed., 2003. Privileged Access: Philosophical Accounts of Self-Knowledge. Hampshire, UK: Ashgate Publishing.

Dretske, F.,2006. The Epistemology of Pain. In: Aydede M.,ed., 2006. Pain: New Essays on its Nature and the Methodology of Its Study. Cambridge, Mass.: MIT Press.

Dowell, J.,2006. Formulating the Thesis of Physicalism: An Introduction. Philosophical Studies, 131, pp. 1-23.

Ellis, R. D. & Newton, N.,1998. Three paradoxes of phenomenal consciousness: Bridging the explanatory gap. Journal of Consciousness Studies, *5*, pp. 419-42.

Everitt, N.,1988. Pain and Perception. Proceedings of the Aristotelian Society, 89, pp. 113-124.

Fields, H. L.,1999. Pain: An Unpleasant Topic. Pain Supplement, 6, pp. 61-69.

Flanagan, O. J.,1985. Consciousness, naturalism and Nagel. Journal of Mind and Behavior, 6, pp. 373-90.

Fleming, N.,1976. The Objectivity of Pain. Mind, 85*(340),* pp. 522-541.

Fodor, J.,1987. Psychosemantics: The Problem of Meaning in the Philosophy of Mind. Cambridge, Massachusetts: MIT Press.

Foltz, E. L. and White E. W.,1962a. The Role of Rostral Cingulumonotomy in 'Pain' Relief. International Journal of Neurology, 6, pp. 353-373.

Foltz, E. L. and White, L. E.,1962b. Pain'Relief' by Frontal Cingulotomy. Journal of Neurosurgery, 19, pp. 89-100.

Foss, J. E., 1989. On the logic of what it is like to be a conscious subject. Australasian Journal of Philosophy, 67, pp. 305-320.

Foss, J. E.,1993. Subjectivity, objectivity, and Nagel on consciousness. Dialogue, 32, pp. 725-36.

Foster, J.,2000. The Nature of Perception. Oxford: Oxford University Press.

Francescotti, R. M.,1993. Subjective experience and points of view. Journal of Philosophical Research, 18, pp. 25-36.

Freeman, W. and Watts, J. W.,1950. Psychosurgery, in the treatment of mental disorders and intractable pain. Springfield, IL: Charles.

Freeman, W., Watts, J. W. and Hunt, T.,1942. Psychosurgery; intelligence, emotion and social behavior following prefrontal lobotomy for mental disorders. Springfield, IL: C. C. Thomas.

Freeman, W. and Wattz, J. W.,1946. Pain of Organic Disease Relieved by Prefrontal Lobotomy. Proceedings of the Royal Academy of Medicine, 39, pp. 44-447.

Fodor, J.,1974. Special Sciences (or: The Disunity of Science as a Working Hypothesis). Synthese, 28, pp. 97-115.

García-Carpintero, M. and Macià, J.,eds.,2006. Two-Dimensional Semantics. Oxford: Oxford University Press.

Gertler, B.,2002. Explanatory reduction, conceptual analysis, and conceivability arguments about the mind. Nous, 36, pp. 22-49.

Gertler, B.,2001. The explanatory gap is not an illusion: A reply to Michael Tye. Mind, 110, pp. 689-694.

Gois, I.,2001. Understanding consciousness. Disputatio, 10, pp. 3-21.

Gracely, R. H., R. Dubner and McGrath, P. A., 1979. Narcotic Analgesia: Fentanyl Reduces the Intensity but not the Unpleasantness of Painful Tooth Pulp Sensations. Science, 203, pp. 1261-63.

Graham, G. and Stephens, G. L., 1985. Are Qualia a Pain in the Neck for Functionalists? American Philosophical Quarterly, 22, pp. 73-80.

Grahek, N.,1991. Objective and Subjective Aspects of Pain. Philosophical Psychology, 4, pp. 249-266.

Grahek, N.,2007. Feeling Pain and Being in Pain. Cambridge, Massachusetts: MIT Press.

Grice, H. P.,1962. Some Remarks About the Senses. In: Butler R. J., ed., 1962. Analytical Philosophy. Oxford, UK: Blackwell.

Guirguis, M. M.,1998. Robotoid Arthritis or How Humans Feel Pain. Philosophical Writings, 7, pp. 3-12.

Gustafson, D.,1998. Pain, Qualia and the Explanatory Gap. Philosophical Psychology, 11, pp. 371-387.

Gustafson, D.,2006. Categorizing Pain. In: Aydede M. ed., 2006. Pain: New Essays on its Nature and the Methodology of Its Study. Cambridge, Mass.: MIT Press.

Hacker, P. M. S.,2002. Is there anything it is like to be a bat? Philosophy, 77, pp. 157-174.

Haksar, V. 1981. Nagel on Subjective and objective. Inquiry, 24, pp. 105-21.

Hanfling, O.,2003. Wittgenstein and the problem of consciousness. Think, 3, pp. 99-106 .

Hanna, P. 1990. Must thinking bats be conscious? Philosophical Investigations, 13, pp. 350-55.

Hardin, L. C.,1987. Qualia and Materialism: Closing the Explanatory Gap. Philosophy and Phenomenological Research, 48, pp. 281-298.

Hardin, C. L.,1992. Physiology, phenomenology, and Spinoza's true colors. In: Beckermann, A., Flohr, H. and Kim, J. eds., 1992. Emergence or Reduction?: Prospects for Nonreductive Physicalism. Berlin-New York: De Gruyter.

Hardcastle, V. G.,1997. When a Pain is Not. The Journal of Philosophy, 94*(8),* pp. 381-409.

Hardcastle, V. G.,1999. The Myth of Pain. Cambridge, Massachusetts: MIT Press.

Hardy, J. D., Wolff, H. J. and Goodell H., 1952. Pain Sensations and Reactions. Baltimore, MD: Williams and Wilkins.

Harnad, S.,1994. Why and how we are not zombies. Journal of Consciousness Studies, 1, pp. 164-67.

Harman, G.,1990. The Intrinsic Quality of Experience. Philosophical Perspectives 4, pp.31-52.

Hempel, C.,1969. Reduction: Ontological and Linguistic Facets. In: Morgenbesser, S., Suppes, P. and White, M.,eds., 1969. Philosophy, Science, and Method: Essays in Honor of Ernest Nagel. New York: St. Martin's Press.

Hiley, D. R. 1978. Materialism and the inner life. Southern Journal of Philosophy, 16, pp. 61-70.

Hill, C. S.,1977. Of bats, brains, and minds. Philosophy and Phenomenological Research, 38, pp. 100-106.

Hill, C. S.,1997. Imaginability, Conceivability, Possibility, and the Mind-Body Problem. Philosophical Studies, 87 *(1),* pp. 61–85.

Hill, C. and McLaughlin, B., 1999. There Are Fewer Things in Reality that Are Dreamt of in Chalmers' Philosophy. Philosophy and Phenomenological Research, 59, pp. 445-454.

Hill, C. S.,2004. Ouch! An Essay on Pain. In: Gennaro J., ed., 2004. Higher Order Theories of Consciousness. Amsterdam: John Benjamins.

Hill, C.,2006. Ow! The Paradox of Pain. In: Aydede, M., ed., 2006. Pain: New Essays on Its Nature and the Methodology of Its Study. Cambridge. Mass.: MIT Press.

Holborow, L. C.,1969. Against Projecting Pains. Analysis, 29, pp. 105-108.

Hurt, R. W. and Ballantyne H. T.,1974. Stereotactic Anterior Cingulate Lesions for Persistent Pain: A Report on 68 Cases. Clinical Neurosurgery, 21, pp. 334-351.

IASP, 1986. Pain Terms: A List with Definitions and Notes on Pain. Pain, 3(Suppl).

IASP-Task-Force-On-Taxonomy,1994. IASP Pain Terminology. In: Merskey H. and Bogduk N., eds., 1994. Classification of Chronic Pain: Descriptions of Chronic Pain Syndromes and Definitions of Pain Terms. Seattle, WA: IASP Press.

Jackson, F.,1975. On the Adverbial Analysis of Visual Experience, Metaphilosophy, 6, pp. 127-135.

Jackson, F.,1977. Perception: A Representative Theory. Cambridge: Cambridge University Press.

Jackson, F.,1982. Epiphenomenal Qualia, Philosophical Quarterly, 32 *(127),* pp. 127-136.

Jackson, F.,1986. What Mary Didn't Know. Journal of Philosophy, 83 *(5),* pp. 291-295.

Jackson, F.,1994. Finding the Mind in the Natural World. In: Casati, R., Smith, B. and White, G.,eds., 1994. Philosophy and the Cognitive Sciences. Vienna: Holder-Pichler-Tempsky.

Jackson, F.,1998a. From Metaphysics to Ethics: A Defence of Conceptual Analysis. New York: Oxford University Press.

Jackson, F.,1998b. Reference and Description Revisited. Nous, 32, pp. 201-218.

Jackson, F.,2003. Mind and Illusion. In: O'Hear A.,ed., 2003. Minds and Persons (Royal Institute of Philosophy Supplements 53). Cambridge: Cambridge University Press.

Kaufman, R.,1985. Is the Concept of Pain Incoherent? Southern Journal of Philosophy, 23 *(2),* pp. 279–283.

Kekes, J.,1977. Physicalism and subjectivity. Philosophy and Phenomenological Research, 37, pp. 533-6.

Kim, J.,1993. Supervenience and Mind. Cambridge: Cambridge University Press.

Kim, J.,1998. Mind in a Natural World. Cambridge, MA: MIT Press.

Kim, J.,2005. Physicalism, or Something Near Enough. Princeton: Princeton University Press.

Kirk, R.,1974. Zombies vs. Materialists. Proceedings of the Aristotelian Society, Supp., 48, pp. 135-52.

Kirk, R., 2005. Zombies and Consciousness. Oxford: Oxford University Press.

Kraut, R.,1982. Sensory States and Sensory Objects. Nous, 16*(2),* pp. 277-293.

Kripke, S.,1980. Naming and Necessity. Cambridge Massachusetts: Harvard University Press.

LeDoux, J.,1996. The Emotional Brain. The mysterious Underpinnings of Emotional Life. New York : Touchstone.

Levin, J.,1991. Analytic Functionalism and the Reduction of Phenomenal States. Philosophical Studies, 61, pp. 211-238.

Levine, J.,1983. Materialism and Qualia: The Explanatory Gap. Pacific Philosophical Quarterly, 64, pp. 354-361.

Levine, J.,1993.On Leaving Out What It's Like. In: Davies M. and Humphreys G., eds., 1993. Consciousness: Psychological and Philosophical Essays, Oxford: Blackwell.

Levine, J.,1998. Conceivability and the metaphysics of mind. Nous, 32*(4),* pp. 449-480.

Levine, J.,1999. Conceivability, Identity, and the Explanatory Gap. In: Hameroff, S. R. and Chalmers, D.J. eds., 1999. Toward a Science of Consciousness III: The Third Tucson Discussions and Debates. Cambridge, MA: MIT Press.

Levine, J.,2001. Purple Haze: The Puzzle of Conscious Experience. Cambridge, MA: MIT Press.

Lewis, D.,1983a. Mad pain and Martian pain. In: Block, N., ed. 1983. Readings in philosophy of psychology, Volume 1. Cambridge, Massachusetts: Harvard University Press.

Lewis, D.,1983b. Postscript to "Mad pain and Martian pain". In: Lewis, D. ed., 1983. Philosophical Papers, 1. New York Oxford: Oxford University Press.

Lewis, D.,1983c. Extrinsic Properties. Philosophical Studies, 44, pp. 197-200.

Lewis, D.,1994. David Lewis: Reduction of Mind. In: Guttenplan, S. ed. 1994. A Companion to the Philosophy of Mind. Oxford: Blackwell.

Langsam, H.,1995. Why Pains are Mental Objects. The Journal of Philosophy, 92*(6),* pp. 303-313.

Lycan, W. G.,1987a. Consciousness. Cambridge, Massachusetts: MIT Press.

Lycan, W. G.,1987b. Phenomenal Objects: A Backhanded Defense. In: Tomberlin, J. E., ed., 1987. Philosophical Perspectives. Atascadero: Ridgeview.

Lycan, W. G.,1987c. Subjectivity. In: Lycan, W. G.,1987a. Consciousness. Cambridge, Massachusetts: MIT Press.

Lycan, W. G.,1990. What is the "subjectivity" of the mental? Philosophical Perspectives, 4, pp. 109-130.

Loar, B.,1990. Phenomenal States. Philosophical Perspectives, 4, pp. 81-108.

Loar, B.,1997. Phenomenal States (Revised Edition). In: N. Block, O. Flanagan and G. Güzeldere . eds., 1997. The Nature of Consciousness: Philosophical Debates. Cambridge, MA: MIT Press.

Loar, B.,1999. David Chalmers' The Conscious Mind. Philosophy and Phenomenological Research, 59, pp. 465-472.

Loar, B., 2002. Phenomenal states. In: Chalmers, D.,ed., 2002c. Philosophy of Mind: Classical and Contemporary Readings. New York: Oxford University Press.

Loar, B.,2003. Qualia, Properties, Modality, Philosophical Issues, 13 (*1*), pp. 113-129.

Ludlow, P., Nagasawa, Y. and Stoljar, D.,eds.,2004. There's Something about Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument. Cambridge, MA: MIT Press.

Lycan, W. G.,1996a. Layered Perceptual Representation. Philosophical Issues, 7, *Perception,* pp. 81-100.

Lycan, W. G.,1996b. Consciousness and experience. Cambridge, MA: MIT Press.

Macdonald, C.,2004. Mary Meets Molyneux: The Explanatory Gap and the Individuation of Phenomenal Concepts. Nous, 38, pp. 503–524.

Malatesti, L.,2004. Knowing what it is like and knowing how. In: Peruzzi, A. ed., 2004. Mind and Causality. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Malcolm, N.,1988. Subjectivity. Philosophy, 63, pp. 147-60.

Maloney, J. C.,1986. About being a bat. Australasian Journal of Philosophy, 64, pp. 26-49.

Mandik, P.,2001.Mental representation and the subjectivity of consciousness. Philosophical Psychology, 14, pp. 179-202.

Manson, N.,2002. Consciousness-dependence and the explanatory gap. Inquiry, 45, pp. 521-540.

Marcus, E.,2004. Why Zombies are Inconceivable. Australasian Journal of Philosophy, 82, pp. 477-490.

Margolis, J.,1976. Pain and Perception. International Studies in Philosophy, 8, pp. 3-12.

Maund, B.,2003. Perception. Chesham: Acumen.

Maund, B.,2006. Tye on Pain and Representational Content. In: Aydede, M. ed., 2006. Pain: New Essays on Its Nature and the Methodology of Its Study. Cambridge, Mass.: MIT Press.

Mayberry, T. C.,1978. The Perceptual Theory of Pain. Philosophical Investigations, 1, pp. 31-40.

Mayberry, T. C.,1979. The Perceptual Theory of Pain: Another Look. Philosophical Investigations, 2, pp. 53-55.

McClamrock, R.,1992. Irreducibility and subjectivity. Philosophical Studies, 67, pp. 177-92.

McCulloch, G.,1988. What it is like. Philosophical Quarterly, 38, pp. 1-19.

McLaughlin, B.,1995. Varieties of Supervenience. In: Savellos E. and Yalçin U., eds., 1995. Supervenience: New Essays. New York: Cambridge University Press.

McMullen, C., 1985. `Knowing what it's like' and the essential indexical. Philosophical Studies, 48, pp. 211-33.

McKenzie, J. C.,1968. The Externalization of Pains. Analysis, 28: 189-193.

McMurray, G. A.,1955. Congenital insensitivity to pain and its implications for motivational theory. Canadian Journal of Psychology, 9*(2),* pp. 121-131.

McMurray, G. A.,1975. Theories of pain and congenital universal insensitivity to pain. Canadian Journal of Psychology, 29(*4*), pp. 302-315.

Mellor, D. H.,1993. Nothing like experience. Proceedings of the Aristotelian Society, 63, pp. 1-16.

Melnyk, A.,2003. Physicalism. In: Warfield T. and Stich S .,eds. 2003., The Blackwell guide to Philosophy of mind. Oxford: Blackwell.

Melzack, R.,1973. The Puzzle of Pain. New York: Basic Books.

Melzack, R. and Casey, K. L.,1968. Sensory, Motivational, and Central Control Determinants of Pain: A New Conceptual Model. In: Kenshalo D., ed., 1968. The Skin Senses. Springfield: Charles C. Thomas.

Melzack, R. and Wall, P. D.,1965. Pain Mechanisms: A New Theory. Science, 150, pp. 971-979.

Melzack, R. and Torgerson, W.S.,1971. On the language of pain. Anesthesiology, 34 (*1*), pp. 50-59.

Melzack, R. and Wall, P. D.,1988. The Challenge of Pain. London, England: New York, NY: Penguin Books.

Millikan, R. G.,1984. Language, Thought, and Other Biological Categories: New Foundations for Realism. Cambridge, MA: MIT Press.

Montero, B.,2002. The epistemic/ontic divide. Philosophy and Phenomenological Research, 66, pp. 404-418.

Moore, G. E.,1939. Proof of an External World. Proceedings of the British Academy, 25, pp. 273-300.

Mounce, H. O.,1992. On Nagel and consciousness. Philosophical Investigations, 15, pp. 178-84.

Muscari, P.,1985. The subjective character of experience. Journal of Mind and Behavior, 6, pp. 577-97.

Muscari, P.,1987. The status of humans in Nagel's phenomenology. Philosophical Forum, 19, pp. 23-33.

Musacchio, J. M.,2002. Dissolving the explanatory gap: Neurobiological differences between phenomenal and propositional knowledge. Brain and Mind, 3, pp. 331-365.

Nagel, T.,1974. What is it Like to Be a Bat? Philosophical Review, 83 (*4*), pp. 435-450.

Nagel, T.,1979. Subjective and objective. In: Nagel, T.,1979. Mortal Questions. Cambridge: Cambridge University Press.

Nagel, T.,1986. The View From Nowhere. Oxford: Oxford University Press.

Nagasawa, Y.,2003. Thomas versus Thomas: A new approach to Nagel's bat argument. Inquiry, 46, pp. 377-395.

Nelkin, N.,1987. What is it like to be a person? Mind and Language*,* 2, pp. 220-41.

Nemirow, L.,1980. Review of Nagel's Mortal Questions. Philosophical Review, 89, pp. 473-7.

Nemirow, L.,1990. Physicalism and the cognitive role of acquaintance. In: Lycan, W., ed., 1990. Mind and Cognition. Oxford: Blackwell.

Newton, N.,1989. On Viewing Pain As a Secondary Quality. Nous, 23(*5*), pp. 569-598.

Papineau, D.,1987. Reality and Representation. London: Basil Blackwell.

Papineau, D.,1993. Physicalism, consciousness and the antipathetic fallacy. Australasian Journal of Philosophy, 71, pp.169–183.

Papineau, D.,1998. Mind the gap. Philosophical Perspectives, 12, pp. 373-88.

Papineau, D.,2002. Thinking about Consciousness. Oxford: Oxford University Press.

Papineau, D.,2007. Phenomenal and Perceptual Concepts. In: Alter T. and Walter S.,eds., 2007. Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism. Oxford: Oxford University Press.

Pauen, M.,1998. Is there an empirical answer to the explanatory gap argument? Commentary on Taylor, J. G., Cortical Activity and the Explanatory Gap. Consciousness and Cognition, 7, pp. 202-205.

Pauen, M. ,1999. Phenomenal Experience and Science: Separated by a “brick wall”? Commentary on: Palmer, S. E., Color, Consciousness, and the Isomorphism Constraint.The Behavioral and Brain Sciences, 22, pp. 968-968.

Pauen, M.,2000. Painless Pain: Property-Dualism and the Causal Role of Phenomenal Consciousness. American Philosophical Quarterly, 37, pp. 51-64.

Pauen, M.,2003. Is type identity incompatible with multiple realization? Grazer Philosophische Studien, 65, pp. 37-49.

Pauen, M.2006. Felling Causes. Journal of Consciousness Studies, 13 (*1–2*), pp. 129-152.

Peacocke, C.,1989. What Are Concepts? Midwest Studies In Philosophy, 14 (*1*), pp. 1–28.

Peacocke, C.,1991. The Metaphysics of Concepts. Mind, C (*400*), pp. 525-546.

Perry, J.,2001. Knowledge, Possibility and Consciousness. Cambridge, MA: MIT Press.

Perkins, M.,1983. Sensing the World. Indianapolis, Indiana: Hackett.

Perkins, M.,2006. An Indirectly Realistic, Representational Account of Pain Perception. In Aydede, M. ed., 2006. Pain: New Essays on Its Nature and the Methodology of Its Study. A Bradford Book. Cambridge Mass.: MIT Press.

Pitcher, G. 1969. Mckenzie on Pains. Analysis, 29, pp. 103-105.

Pitcher, G.,1970. Pain Perception. The Philosophical Review, 79(*3*), pp. 368-393.

Pitcher, G.,1971. A Theory of Perception. Princeton, NJ: Princeton University Press.

Pitcher, G.,1978. The Perceptual Theory of Pain: A Response to Thomas Mayberry's, "the Perceptual Theory of Pain". Philosophical Investigations, 1, pp. 44–46.

Place, U. T.,1956. Is Consciousness a Brain Process? British Journal of Psychology, 47, pp. 44-50.

Ploner, M., H. J. Freund and Schnitzler A.,1999. Pain Affect without Pain Sensation in a Patient with a Postcentral Lesion. Pain 81(*1/2),* pp. 211-214.

Price, D. D.,1988. Psychological and Neural Mechanisms of Pain. New York: Raven Press.

Price, D. D.,1999. Psychological Mechanisms of Pain and Analgesia. Seattle: IASP Press.

Price, D. D.,2000. Psychological and Neural Mechanisms of the Affective Dimension of Pain. Science, 288(*9*), pp.1769-72.

Price, D. D.,2002. Central Neural Mechanisms that Interrelate Sensory and Affective Dimensions of Pain. Molecular Interventions, 2, pp. 392-403.

Price, H. H.,1950. Perception. London: Methuen.

Price, M. C.,1996. Should we expect to feel as if we understand consciousness? Journal of Consciousness Studies, 3, pp. 303-12.

Putnam, H.,1967. Psychological Predicates. In: Capitan W. H. and Merrill D. D., eds. Art, Mind and Religion. Pittsburgh: University of Pittsburgh Press.

Putnam, H.,1968. Brains and Behaviour. In: Butler R. ed., 1968. Analytical Philosophy:Second Series. Oxford: Blackwell.

Putnam, H.,1975. The Meaning of 'Meaning'. In: Gunderson, K.,ed., 1975. Language,Mind and Knowledge. Minneapolis: University of Minnesota Press.

Rainville, P., Carrier, B., Hofbauer, R. K., Bushnell M. C., and Duncan G. H.,1999. Dissociation of sensory and affective dimensions of pain using hypnotic modulation. Pain Forum 82(*2*), pp. 159-171.

Rainville, P., Duncan, G. H., Price, D. D., Carrier, B. and Bushnell, M. C., 1997. Pain Affect Encoded in Human Anterior Cingulate but not Somatosensory Cortex. Science, 277(*5328*), pp. 968-971.

Robinson, H.,1982. Matter and Sense: A Critique of Contemporary Materialism. Cambridge, New York: Cambridge University Press.

Robinson, H.,1994. Perception. London; New York, NY: Routledge.

Rubins, J. L. and Friedman E. D.,1948. Asymbolia for Pain. Arch. Neurol. Psychiat., 60, pp. 554-73.

Rosenthal, D.,2010. How to think about metal qualities. 20, *1,* pp. 368–393.

Rudd, A. J.,1999. What it's like and what's really wrong with physicalism: A Wittgensteinian perspective. Journal of Consciousness Studies, 5, pp. 454-63.

Russow, L.,1982. It's not like that to be a bat. Behaviorism, 10, pp. 55-63.

Scheele, M.,2002. Never mind the gap: The explanatory gap as an artifact of naive philosophical argument. Philosophical Psychology, 15 (*3*), pp. 333-342.

Schwitzgebel, E.,2002. Why did we think we dreamed in black and white? Studies in History and Philosophy of Science, 33, pp. 649–660.

Schwitzgebel, E.,2002b. How well do we know our own conscious experiance? Journal of Consciousness Studies, 9 (*5-6*), pp. 35-53.

Seager, W.,2002. Emotional Introspection. Consciousness and Cognition, 11(*4*), pp. 666-687.

Sellars, W.,1963. Science, Perception, and Reality. Atascadero, CA: Ridgeview.

Sellars, W.,1975. The Adverbial Theory of the Objects of Sensation. Metaphilosophy, 6, pp. 144-160.

Sellars, W.,1981. Is consciousness physical? Monist, 64, pp. 66-90.

Shiner, R. A.,1977. Goldman on the non-contingency thesis. Mind, 86, pp. 587-90.

Shoemaker, S.,1982. The inverted spectrum. Journal of Philosophy, 79, pp. 357-81.

Simoni-Wastila, H., 2000. Particularity and consciousness: Wittgenstein and Nagel on privacy, beetles and bats. Philosophy Today, 44, pp. 415-425.

Smart, J. J. C.,1959. Sensations and Brain Processes. Philosophical Review, 68, pp. 141-156.

Spurrett, D. & Papineau, D. 1999. A note on the completeness of physics. Analysis, 59 (*1*), pp. 25-29.

Stalnaker, R.,1978. Assertion. In: Cole P.,ed.1978 Syntax and Semantics: Pragmatics. Vol. 9, New York: Academic Press.

Stalnaker, R.,1996. Varieties of Supervenience. Nous, 30, pp. 221-241.

Stanley, J.,1997. Names and Rigid Designation. In: Hale B. and Wright C.,eds., 1997. A Companion to the Philosophy of Language. Oxford: Blackwell Press.

Stephens, G. L. and Graham, G., 1987. Minding your P's and Q's: Pain and Sensible Qualities. Nous 21(*3*), pp. 395-405.

Sternbach, R. A.,1963. Congenital insensitivity to pain: A critique. Psychological Bulletin, 60(*3*), pp. 252–264.

Strawson, G.,1989. Red and 'red'. Synthese 78 *(2*), pp. 193-232.

Sturgeon, S.,1994. The epistemic basis of subjectivity. Journal of Philosophy, 91, pp. 221-35.

Sturgeon, S.,1998. Physicalism and Overdetermination. Mind, 107 (*426*), pp. 411-432.

Sturgeon, S., 2000. Matters of Mind. London: Routledge.

Stoljar, D.,2000. Physicalism and the Necessary a Posteriori. The Journal of Philosophy, 97 (*1*), pp. 33-54.

Stoljar, D.,2001a. Two Conceptions of the Physical. Philosophy and Phenomenological Research, 62, pp. 253-281.

Stoljar, D.,2005. Physicalism and Phenomenal Concepts. Mind and Language, 20, pp. 469-494.

Stoljar, D.,2006. Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness. Oxford: Oxford University Press.

Stoljar, D.,2009a. The Conceivability Argument and Two Conceptions of the Physical. Nous, 35, pp. 393-413.

Stoljar, D.,2009b. Physicalism, The Stanford Encyclopedia of Philosophy. [internet encyclopedia]

Available at: Stanford Encyclopedia of Philosophy.
http://plato.stanford.edu/entries/physicalism/
[Accessed 09.09. 2009]

Taliaferro, C.,1988. Nagel's vista or taking subjectivity seriously. Southern Journal of Philosophy, 26, pp. 393-401.

Teller, P.,1992. Subjectivity and knowing what it's like. In: Beckermann, A., Flohr, H. and Kim, J. eds., 1992. Emergence or Reduction?: Prospects for Nonreductive Physicalism. Berlin-New York: De Gruyter.

Tilghman, B. R.,1991. What is it like to be an aardvark? Philosophy, 66, pp. 325-38.

Tye, M.,1996. Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind. Cambridge, Massachusetts: MIT Press.

Tye, M.,1999. Phenomenal Consciousness: the Explanatory Gap as a Cognitive Illusion. Mind, 108, pp. 705-725.

Tye, M.,2001. Oh yes it is. Mind, 110, pp. 695-697.

Tye, M., 2003. A theory of phenomenal concepts. In: O'Hear A.,ed., 2003. Minds and Persons. Cambridge: Cambridge University Press.

Tye, M.,1984a. The Adverbial Approach to Visual Experience. The Philosophical Review, 93(*2*), pp. 195-225.

Tye, M.,1984b. Pain and the Adverbial Theory. American Philosophical Quarterly, 21, pp. 319-328.

Tye, M.1997. A Representational Theory of Pains and their Phenomenal Character. In: Block, N., Flanagan O. and Güzeldere, G. eds., 1997. The Nature of Consciousness: Philosophical Debates. Cambridge, Massachusetts: MIT Press.

Tye, M.,2006a. Another Look at Representationalism about Pain. In: Aydede, M. ed., 2006. Pain: New Essays on Its Nature and the Methodology of Its Study. Cambridge: Mass.: MIT Press.

Tye, M.,2006b. In Defense of Representationalism: Reply to Commentaries. In: Aydede M., ed., 2006. Pain: New Essays on Its Nature and the Methodology of Its Study. Cambridge, Mass.: MIT Press.

van Gulick, R.,1985. Physicalism and the subjectivity of the mental. Philosophical Topics, 13, pp. 51-70.

van Gulick, R., 2002. Maps, gaps, and traps. In : Smith Q. and Jokic, A. eds., 2002. Consciousness: New Philosophical Essays. Oxford: Oxford University Press.

Varela, F.,1996. Neurophenomenology : A Methodological Remedy for the Hard Problem. Journal of Consciousness Studies, 3 (*4*), pp. 330-349.

Vesey, G. N. A.,1964a. Armstrong on Bodily Sensations. Philosophy, 39, pp. 177-181.

Vesey, G. N. A.,1964b. Bodily Sensations. Australasian Journal of Philosophy, 42, pp. 232-247.

Villanueva, E., Ed.,1996. Perception. (Philosophical Issues 7). Atascadero, CA: Ridgeview.

Webster, W. R.,2002. A case of mind/brain identity: One small bridge for the explanatory gap. Synthese, 131, pp. 275-287.

Webster, W. R.,2003. Revelation and transparency in colour vision refuted: A case of mind/brain identity and another bridge over the explanatory gap. Synthese, 133, pp. 419-39.

Warfield, T. and Stich, S. eds., 2003. Blackwell Guide to Philosophy of Mind. Oxford: Blackwell.

White, S.,1986. Curse of the Qualia. Synthese 68 (*2*), pp. 333-368.

White, S., 2010. Why the Property Dualism Argument Won't Go Away. In

Bealer G. and Koons R., eds. 2010. The Waning of Materialism: New Essays. Oxford: Oxford University Press.

Wider, K.,1989. Overtones of solipsism in Nagel's `What is it like to be a bat?' and `The view from nowhere'. Philosophy and Phenomenological Research, 49, pp. 481-99.

Wright, E.,1996. What it isn't like. American Philosophical Quarterly, 33, pp. 23-42.

Williamson, T.,2003. Blind Reasoning II: Understanding and Inference. Proceedings of the Aristotelian Society, Supp., 77, pp. 249–293.

Wilson, J., 2006. On Characterizing the Physical. Philosophical Studies ,131, pp. 61-99.

Worley, S.,2003. Conceivability, Possibility and Physicalism. Analysis, 63, pp. 15-23.

Worley, S., 2006. Physicalism and the Via Negativa. Philosophical Studies, 131, pp. 101-126.

Weinstein, E. A., R. L. Kahn and Slate W. H.,1995. Withdrawal, Inattention, and Pain Asymbolia. Archives of Neurology and Psychiatry, 74(*3*), pp. 235-248.

White, J. C. and Sweet W. H., 1969. Pain and the Neurosurgeon: A Forty-Year of Experience. Springfield, IL: Charles C. Thomas.

White, S. L.,1986. Curse of the Qualia. Synthese, 68, pp. 333-367.

Wilkes, K. V.,1977. Physicalism. London, UK: Routledge.