Relocating the Gap:
Physicalism and Phenomenal Concepts

Inauguraldissertation
zur Erlangerung des Grades eines Doktors der Philosophie
im Fachbereich Philosophie-und Geschichtswissenschaften
der Johann Wolfgang Goethe-Universität
zu Frankfurt am Main

vorgelegt von

Julia Cavalcanti Telles de Menezes
aus: Rio de Janeiro

2015

1. Gutachter: Prof. Dr. André Fuhrmann (Frankfurt)
2. Gutachter: Prof. Dr. Wilson Mendonça (Rio de Janeiro)

Contents

Acknowledgments

First of all, I am very grateful to my supervisor, Prof. André Fuhrmann, for his support during my stay in Frankfurt. He made me feel incredibly welcomed since the day of my arrival. He has been very patient, dedicated and generous with his time at all stages of this PhD. So, once again, thank you for the support, for the philosophical discussions and dedication especially during the awful amount of work near the end of the writing process.

I am also grateful to Prof. Wilson Mendonça, who encouraged me since my bachelor years in Rio, who has been a wonderful friend, advisor and has provided exciting philosophical discussions.

Additionally, I wish to thank CAPES and DAAD for their financial support, which enabled my stay in Germany.

I would also like to thank all my friends, many of which are still in Rio, and the new friends I made in Frankfurt, for the so needed moments of leisure. I am also deeply grateful to my parents for their unlimited support. Finally, I am forever in debt to Gustavo for his patience, friendship, endurance, and lastly for making Frankfurt my home.

Introduction

The dominant world-view in the pre-modern era used to favor explanations of certain natural events that involved all kinds of immaterial entities such as gods, angels and magical creatures in general. Back then, the way to explain seizures, for example, was to consider it as a sign of demoniac possession; once, the cycle of the sun was attributed to Apollo and his horse moving across a flat Earth; what is now understood as a mental health issue was once related to curses, spells or something equally spooky. At a certain point in history, a scientific turn took place so that it progressively eliminated such explanations as inadequate accounts of the relevant phenomena. Such magical explanations were replaced by scientific ones. One requirement for the best explanation would be that it should not be easily replaced by other explanations without any loss in explanatory force. The scientific world-view that ties seizures with neurological disfunction, the solar cycle with the earth's movement around the sun and around itself, or even that psychiatric problems should be explained in terms of genetic and environmental influences are far better suited to replace the magical explanations previously available. The scientific turn is what Weber called the phenomenon of 'disenchantment of the world' (*Entzauberung der Welt*). Modern science endorses, as background methodology, the existence of a physicalist world-view, which is established after a change of paradigm in our world, that is to say, after the disenchantment of the world. Clearly, this change in paradigm is not accomplished without resistance and controversy. In the past, many were burned at the cross for suggesting such a radical change of view. Nowadays the reaction is milder, yet the questions are still source of great dispute. Nevertheless, this 'new' physicalist world-view is

still a pre-philosophical view about what there is in the world. [1] The physicalist world-view is sometimes understood as a package of views distinct from the metaphysical thesis of physicalism. It involves, for example, the idea that the methodology employed in natural sciences will provide complete theoretical knowledge of the world and that this way of understanding the world will deliver a final theory of everything. It also claims that this complete theory of everything will be, like natural sciences, objective, hence, it will reduce subjective perspectives to objective vocabulary. The view also involves the idea that all relevant explanation is physically reductive; that physics is general enough to reductively explain events that special sciences already explained, because every event is held to have a physical cause; and it is sometimes added to this the idea of atheism. Nevertheless, it would be a mistake to think that physicalism implies all of these views. For example, atheism is not implied nor implies physicalism, since it is consistent with physicalism and anti-physicalism. On the contrary, physicalism is, at least neutral, regarding many of the views. This is why we should distinguish the physicalist pre-philosophical world-view from the metaphysical thesis of physicalism.

One thing that the physicalist world-view and the metaphysical thesis of physicalism have in common is the strong intuition that physicalism generates a certain tension when confronted with some central features of our everyday life. This monist view is *prima facie* incompatible with features such as: abstraction, intentionality, phenomenality, normativity etc.. If we grant that the tension is real, the core question of physicalism arises of how to accommodate such aspects of our everyday life (e.g. mentality and phenomenality) in this monist, physical world. This will be our starting point in the task of finding a more sophisticated definition of physicalism that is able to include features of mentality and phenomenality. If such definition is not available, the physicalist needs to provide good reasons to discard mentality and phenomenality. There are at least three routes out of this

[1] I follow Stoljar (2010, 2015) in the characterization of the physicalist world-view.

problem. One is to simply deny physicalism and instead recommend a form of dualism. This will resolve the tension, but it will generate new and serious problems as that of explaining non-physical influence in a physical world which in causally closed. Another possibility is to preserve physicalism but to abdicate from mentality by treating it as an illusion that should be eliminated, just like 'magical' explanations were once eliminated. The idea is that the vocabulary that we still use to refer to aspects of mentality can be replaced by a strict physicalist code. The problem with this proposal is that it asks us to give up something too essential to our common life, namely, mental talk. We do want to preserve the mental vocabulary as we do want to preserve mentality. The best way to preserve them is to define a *compromise* version of physicalism that accounts for mentality from a physicalist point of view. This is our central aim. In the course of this work, I want to systematize and contrast positions that aim to solve the *prima facie* incompatibility of mind and matter.

The central aim of this work is the physicalist attempt to respond to two central anti-physicalist arguments, viz. the knowledge argument and the conceivability argument. Both arguments against physicalism confront the subjective character of consciousness with a physicalist world-view. They give rise to the systematization of the common intuition regarding the tension between consciousness and physicalism.

Chapter One will be an attempt to define what is the metaphysical doctrine of physicalism. In this chapter, I will present a definition of physicalism by answering two central questions; how to interpret the physicalist slogan 'everything in physical'; and why physicalism is so plausible. Regarding the plausibility of physicalism, we will analyze a central argument in favor of physicalism and we will defend physicalism from argument against it in the following chapters. Regarding the question about the proper interpretation of physicalism, it will unfold into three different questions: (i) the scope question asks about the restriction of the quantifier in the physicalist slogan; (ii) the base question asks about the definition of 'physical', and (iii) the relation question inquires about the proper interpretation of the relation between 'everything' and 'physical'. Those three question are meant

to provide substance to a preliminary definition of physicalism (under the physicalist slogan) in order to compromise notions that seem to fall out of the definition of physicalism, like intentionality and phenomenality.

Chapter Two will expose and assess the anti-physicalist arguments that confront the subjective character of consciousness and physicalism. The discussion of the conceivability argument will include an exposition of the two-dimensional semantical framework, in particular, David Chalmers' and Frank Jackson's ambitious interpretation of the framework as the relevant interpretation to provide the link between conceivability and possibility required by the conceivability argument. In fact, we will critically discuss the passage from conceivability and possibility. Additionally, I will show that both the knowledge argument and the conceivability argument have the same structure in that both argue an ontological gap from epistemic gaps. Later, physicalist options will be considered. Among them, options that deny the epistemic gap between the physical and the phenomenal. However, the focus of the present essay will be the physicalist response, which is also known as Type-B Materialism, which acknowledge the epistemic gap but thinks one cannot infer an ontological gap. Among others, I will focus on the phenomenal concept strategy. This is the topic of Chapter Three. I will characterize the task of the strategy, which is to mobilize the *sui generis* character of phenomenal concepts to argue that physicalism can be a posteriori. The most promising physicalist response to the anti-physicalist argument is to show that physicalism can be a posteriori. This should block the anti-physicalist conclusion of the arguments considered if we show that they only threaten a priori physicalism. This will be done with help of phenomenal concepts. Finally, in Chapter Four I will critically assess the phenomenal concept strategy previously discussed by considering three powerful objections. I will respond to each of them and conclude that, although the phenomenal concept strategy has some shortcomings, the objections available are not powerful enough to undermine the strategy. The phenomenal concept strategy is still the most promising physicalist defense against anti-physicalist arguments about consciousness.

Chapter 1

Setting the stage: what is Physicalism?

1 Defining Physicalism

Physicalism is roughly the metaphysical thesis that claims that the world is fundamentally physical. The term 'physicalism' was first introduced by Carnap and Neurath to designate instead a semantic thesis: every sentence describing the mental can be translated into sentences in a physical vocabulary. So originally, 'physicalism' was a semantic thesis, whereas 'materialism' was the term used to designate a metaphysical thesis, that is, a thesis about the nature of the world. Materialism was taken to be the doctrine which claimed that everything is matter, whereas the notion of 'matter' is historically understood as something that is extended, located in space and time. An old-school materialist is someone who claims that everything is matter in this sense. However, modern physics renders this view of matter as false by acknowledging the existence of all sorts of physical entities which have no mass or are not extended in space. For the old-school materialists, such physical entities would be, per definition, immaterial. Other philosophers do not see this as a problem, in fact they prefer to stick with the traditional way of characterizing the monist metaphysical thesis, regardless of the role played by progress in physical sciences. Be that as it may, the role that physical sciences play is one among many other reasons to prefer the term

'physicalism' over 'materialism' as the expression that designates the metaphysical claim about what there is in the world. It is important to notice that, while some philosophers prefer to name the monist metaphysical thesis in question 'physicalism' and others prefer 'materialism', there is still a third group, which chooses to talk in terms of 'naturalism' in an attempt to include other natural sciences besides Physics, like Biology, for instance. From this point on I choose to use 'Physicalism' to refer to the metaphysical claim in question[1].

This chapter consists in an attempt to find a more sophisticated definition of physicalism than that provided by the slogan 'everything is physical'. However, our starting point is this slogan. Our strategy for defining physicalism will be by distinguishing between two questions[2]:

> **(i) The interpretation question**: what does it mean to say that everything is physical?
>
> **(ii) The truth question**: is it true that everything is physical?

Of course, the question concerning the truth of some thesis depends on the answer we will provide to the question about the meaning of that thesis. We shall begin with the first question. Regarding the second question about the truth of physicalism, we will present and assess one popular argument for the plausibility of physicalism and the advantages it holds over dualists theories at the end of the present chapter. After that we will turn to the negative side of the truth question, that is, to analyze arguments against physicalism. Our aim is to assess whether there is a reasonable definition of physicalism that overcomes the challenge that anti-physicalist arguments pose.

---

[1]The way the debate is formulated in contemporary philosophy of mind, these terms are used interchangeably, hence they will appear in quotations in this dissertation.

[2]Here I follow Stoljar's suggestion (2009, 2010) of systematizing the discussion about physicalism.

2 The interpretation question: what does 'everything is physical' mean?

The question about how we should understand the thesis of physicalism unfolds into three different questions:

(a) Scope question: What does it mean to say that *everything* is physical?

(b) Base question: What do we mean by *physical*?

(c) Relation question: What is the relation between *everything* and *physical*?

The scope question leaves the specification of the conception of 'physical' open and inquires about the domain of 'everything' satisfying the condition of being physical, whereas the base question inquires about the conception of 'physical', clearly central to physicalist theories. The relation question asks about the connection between the base and the scope, that is, the relation between physical and everything. Is it identity? Is it supervenience? In the first section we will deal with the interpretation question. At the end of the chapter we will sketch the truth question focusing on one argument for physicalism namely, the causal argument. An attempt to answer the truth question in the negative way (a physicalist defense from epistemic arguments) will be the focus of the remaining chapters.

2.1 The scope question

What is it for *everything* to be physical? Although 'everything' works as universal quantifier, it certainly does not quantify over absolutely everything in the world like in the statement 'Everything is identical to itself.' To grant that 'everything' in our slogan has an unrestricted scope is to assume that *absolutely* everything (concrete, abstract, property, particular etc) is physical. That could hardly be true.

The unrestricted quantification is problematic especially when we consider abstract entities like numbers. Numbers are not physical in the sense that chairs are, for instance. Just as institutions like universities or a court

house are also not physical in the way that paradigmatic physical objects are. It is very hard to ascribe physicality to such complex entities, or, at least, it is highly implausible that the truth of physicalism depends on the truth of numbers and universities being physical. Some physicalists (who are also nominalists) may want to endorse the unrestricted quantification. They argue that since numbers cannot be physical, they do not exist. The problem with this suggestion is that one would have to erase from our ontology many other abstract entities such as complex entities. For this reason, some physicalists (who are not nominalists) think it is best to restrict the quantifier's scope. So the scope question asks about the domain of the quantification operator: which things are properly physical?

One possibility is to restrict the domain to *concrete particulars*, where particulars are defined in opposition to properties. Properties are instantiated in particulars objects, whereas particulars are not instantiated at all. We can put it this way: things are particulars, whereas properties should be features of such things. Concrete entities are defined in opposition to abstract entities: concrete items are extended in time and space, whereas abstract things are not. Now, by restricting the quantification to concrete particulars we would have the following definition:

(1) *Concrete Particulars*: physicalism is true if and only if every concrete particular is physical. (Stoljar 2010: 33)

The problem with restricting physicalism to concrete particulars is that this definition overgenerates: it allows the inclusion of substance dualism in the scope of physicalism. A soul, for example, is usually defined as an unextended substance that could coexist with physical substances in the world. Restricting the domain of physicalism to concrete particulars does not prevent the inclusion of such paradigmatic non-physical entities. This would make physicalism true in a world populated with souls, which is not quite acceptable. Hence, restricting the quantificational operator of physicalism to concrete particulars overgenerates, it declares physical some properties which should be excluded.

Now what if we restrict our domain to properties? Then we would have something like this:

(2) *Properties*: Physicalism is true if and only if every property is physical. (Stoljar, 2010: 32)

This way of defining physicalism seems to avoid any kind of dualism. Since restricting the scope of physicalism to properties is inconsistent with property dualism, which is the claim that, although there is only one substance, which is physical, that physical substance might instantiate non-physical properties. And the substance dualist would also be ruled out, since, to think about non-physical substance is to think about it as instantiating non-physical properties. This is inconsistent with restricting physicalism to properties. Notwithstanding, this way of defining physicalism would make uninstantiated properties part of the physical realm. Some physicalists believe in the existence of uninstantiated properties. For example, 'being a perfect circle' is an uninstantiated property, since there are things that are circular, but not perfectly circular. Also the property of being the new professor of Ethics in 2010 at Rio de Janeiro is a property that could have being instantiated, but, as the selection of Professors failed in 2010, that remains as an uninstantiated property. Some philosophers believe that uninstantiated properties exist. Hence, if there are no souls, soul is an uninstantiated property, like the property of being a perfect circle. This property of being a soul would not be physical, but it would, nevertheless, according to the definition given above, be compatible with our definition of physicalism (2). This would make physicalism compatible with property dualism, still the definition of physicalism overgenerates again. It declares uninstantiated properties as part of the physical realm, which should be excluded. A slight modification would take care of the problem: it suffices that we restrict physicalism to *instantiated* properties:

(3) *Instantiated properties*: Physicalism is true if and only if every instantiated property is physical. (Stoljar 2010: 33)

The formulation of physicalism in (3) avoids the problem of compatibility of property dualism with the existence of uninstantiated physical properties. Of course, we still need to specify the other parameters for our complete definition of physicalism.

2.2 The base question: conception of 'physical'

Perhaps, the most difficult problem, one would think, when engaging in the task of defining physicalism, is to define the embedded notion of 'physical'. The dominant conception of *physical* in the literature in the philosophy of mind ties the notion of instantiated physical property to an authoritative role of physical sciences. However, there are also other options to consider: to define physical properties in terms of the properties instantiated in paradigmatic physical objects, or in terms of methods distinctive of the natural sciences, or by contrasting them with paradigmatic non-physical properties such as mental properties, spirits, souls, etc.. We will take some time to explore a few of the possibilities to define 'physical property'.

2.2.1 Object-based definition

> *Object-based definition*:
> A property is physical if and only if it is a property instantiated in *paradigmatic physical objects*.

This possibility consists in taking the classical route to define physical property in terms of paradigmatic physical objects. Of course, we are left with the task of defining paradigmatic physical objects. If we consider, for a moment, what is the classical view of physical object we presumably arrive at a cluster of features like having certain form and size—being extended—and being located in space and time. An intuitive physical object must at least obey these criteria:

> A thing is a material object if it occupies space and endures through time and can move about in space (literally move about, unlike a shadow or a wave or a reflection) and has a surface and has a mass and is made of certain stuff or stuffs. Or, at any rate, to the extent that one was reluctant to say of something that it had various of these features, to that extent one would be reluctant to describe it as a material object (van Inwagen 1990: 17).

However, modern Physics has established that not all physical objects can be measured by size and that not all physical objects have mass and extension.

On that account, some physical objects are indeed located in space, but they lack mass. Modern physics renders the object-based conception of physical properties too naive, thus inapt to define physical properties. According to the object-based definition, many properties recognized by modern physics came out as non-physical, like being an electron, for example, a particle that has mass but is not extended. Since one of the features of paradigmatic physical objects is extendedness, an electron would be, by definition, a non-physical object.

Another problem with this definition of physicalism is that it defines 'physical property' in a way that is compatible with property dualism. Property dualism is the claim that although there is only one kind of substance, viz. the physical substance, which can be the bearer of physical and non-physical properties. Mental properties are instantiated by the physical substance, therefore, a physical property cannot be simply defined as a property of physical objects, since mental properties would then be, by definition, physical properties. Of course, if physicalism is true, mental properties *are* physical properties. The point is only that the result cannot be arrived at by sheer definition.

An additional objection often raised against the object-based conception of a physical property is called the problem of panpsychism. Roughly, panpsychism is the thesis that all physical objects are conscious beings as well as all conscious properties are physical. Here is a passage where Lewis (1983) describes panpsychism:

> It is often noted that psychophysical identity is a two-way street: if all mental properties are physical, then some physical properties are mental. But perhaps not just some but *all* physical properties might be mental as well; and indeed every property of anything might by at once physical and mental. (Lewis 1983: 362)

This extravagant idea is consistent with the truth of the object-based conception of physicalism, though its implausibility is quite clear: it is simply strange to ascribe consciousness to simple physical objects like sofas or rocks. Though counter-intuitive, panpsychism is consistent with the truth of

physicalism defined via reference to the object-based conception of physical property. The problem is similar to that concerning property dualism and the object-based definition of physicalism: if what defines a physical property is that it can be instantiated by a paradigmatic physical object; then being conscious is both a physical property and a mental property. This strategy overgenerates again, it includes as physical entities properties we should exclude.

2.2.2 Theory-based definition

To bypass the problems that arise with the object-based conception of the physical, an alternative definition of physical property is advanced by many philosophers: by 'physical property' we should mean what is within the range of the language of the physical sciences. Thus, physical sciences will play an authoritative role in defining the physical.

> *Theory-based conception*:
> P is a physical property if and only if P is expressed by a predicate of a true physical theory.

Granting physical theories with the authority to determinate what a physical property is, is perhaps the most popular definition of 'physical'. The idea is that we defer to physicists regarding the meaning of 'physical'. According to the theory-based conception, statements usually made by physicists to explain some physical phenomena fixes the reference of 'physical property'. In this sense, the language of physics determines what is the physical. This sort of definition, although popular, raises serious problems of its own, many of them concerning the notion of physical sciences itself. What do we mean by 'physical sciences'? Hempel is the first to launch the issue:

> The language of *what* physics is meant? Surely not that of, say, 18th century physics; for it contains terms like 'caloric fluid', whose use is governed by theoretical assumptions now false. Nor can the language of contemporary physics claim the role of unitary language, since it will no doubt undergo further changes, too. The thesis of physicalism would seem to require a language

> in which a true story of all physical phenomena can be formulated. But it is quite unclear what is to be understood here by a physical phenomenon, especially in the context of a doctrine that has taken a determinedly linguistic turn. (Hempel 1980: 195)

Hempel's observations point to issues that are raised when we rely solely on the language of physical sciences to tell us what 'physical' means. First, choosing the language of a *particular theory* in physics does not allow us to capture the real spirit of physicalism, which is to claim a *general* thesis about the world. Many things that might be paradigmatic cases of physical objects or properties may fail to satisfy the condition of being physical in such a narrow sense (i.e. with regards to one particular physical theory).

What motivates the theory-based conception is the old idea of the 'unity of science' in which all theories will eventually be reduced to physics forming a complete unified science of everything, able to derive all scientific laws from one 'ever more adequate grand scheme'. Science's continuous change is presented as evidence for the implausibility of the so-called unity of science. Even within the physical sciences we have an amazing diversity of theoretical entities, properties and facts which require different methods of investigation. It is, therefore, quite implausible to think of unified science that integrates both astrophysics and genetics. Maybe this requirement of a unified language of science may be weakened so as to accommodate part of what we are looking for. But the main issue remains, that is the dilemma concerning the kind of physics presupposed in the attempted theory-based characterization of physicalism. Hempel objects that any theory-based definition of physicalism will be either trivially true or false. Is it present-day physics that holds this authoritative role? Or a future, complete physics? We know that current physics is subject to continuous change, since there is always the possibility of making progress by discovering new physical properties. If 'physical property' is defined by present-day physics, then properties discovered only by future physics would be, by definition, non-physical. So physicalism would be false. If we have in mind a future, complete physics, that is, a physical theory that explains everything, then genuine mental properties may have to be included in this final physics, making physicalism

trivial. In sum, the first horn of the dilemma says that if physicalism is defined through present-day physics, then it is false. The second horn of the dilemma says that if physicalism is defined through future physics, then it is trivially true. Hence there is no possibility of coming up with a clear concept of physical that relies solely on the authoritative role of physical sciences. Thus, an adequate and non-trivial question of physicalism cannot even be formulated given the theory-based conception.

Hempel's dilemma is formulated as an objection to the general idea of physicalism. The dilemma is designed to yield the conclusion that the question of physicalism does not even make sense, for we cannot define a clear conception of the physical. Of course, one can avoid this objection by following another course in the task of defining the physical. What the dilemma shows is a problem within the theory-based definition for physicalism, not physicalism itself. These remarks lead us to briefly glance at alternatives for defining 'physical property'.

2.2.3 Method-based conception

An alternative to choosing between future or present-day physics would be to consider the possibility of defining physicalism by referring to the methodology of physical sciences: what if 'physical property' is defined by the language used in any science that applies the methodology of physical sciences? In this case, what would determine the meaning of 'physical property' is not physical sciences per se, but its methodology. Nevertheless, this approach is also subject to a dilemma similar to that posed by Hempel. Considering that methods in the physical sciences change over time, we might want to ask: when we refer to the method of physical sciences are we referring to present-day physical science or future physical science? We face the same dilemma as for the theory-based conception. In particular, if we fix methods of those presently adopted, we exclude items that should be ultimately recognized as physical.

2.2.4 *Via Negativa*

Yet another way to define physical in the context of the mind-body debate is to provide negative, contrastive definitions by referring to paradigmatic non-physical things: mentality, consciousness etc.. We might thus arrive at a list of mental properties and the like. The *via negativa* approach would look something like this:

> *Via Negativa*:
> F is a physical property if and only if F is a non-mental property.

The major problem with this view is that it would imply eliminativism about mental properties. If a physical property is defined in terms of what is a paradigmatic non-physical property, such as mentality and phenomenality, there could not be a way of identifying physical properties and mental properties, since they would be, by definition, distinct. Consider that 'pain' is a paradigmatic mental state. Hence, the properties of 'pain' are by definition non-physical, since what defines physical is the fact that it is non-mental. If this is so, then we cannot even begin to make sense of the identification 'pain is stimulation of c-fibers', since 'pain' is non-physical and 'stimulation of c-fibers' is physical. Thus, the *via negativa* definition renders physicalism as incoherent.

So far we have seen that all extant attempts to define 'physical' fail. Each of them either overgenerates: they include properties that should be excluded from the physical realm or undergenerates: they exclude properties that should be included in the physical realm. The object-based conception is simply too naive to be taken serious, because it is based in 'commonsense physics' which is basically Newtonian mechanics applied to the megascopic world. Since there are physical properties that do not fall under the conception paradigmatic physical objects, the object-based definition of physicalism fail. Even if we overlook this first shortcoming, we are still left with a conception of physical that makes physicalism compatible with panpsychism and property dualism, since we might have mental properties figuring as properties of paradigmatic physical objects. The failure of this classical route to defining 'physical' leads to the search for alternatives.

The most popular conception of 'physical' involves the authoritative role of physical sciences. Physics has complete authority in determining what, after all, is physical. This is initially a very attractive position since it relies on an important principle of the physicalist world view, i.e. that the body of physical sciences should be a complete doctrine. However, the theory-based conception of a physical property is susceptible to Hempel's dilemma involving the conception of physical sciences we are presupposing in this definition: present-day physics makes physicalism false and future physics makes physicalism trivial. This objection is designed to yield drastic results for physicalism; since there is no coherent conception of the physical, physicalism cannot even be formulated. We have also explored the prospects of using the methodology of physical sciences as the central feature of the physical, but that is subject to a variant of Hempel's dilemma. Finally, the *via negativa* which defines physical properties by contrasting it with mental properties has showed to be inadequate, for it has the consequence that the idea of physicalism is incoherent.

2.2.5 Revisiting the theory-based conception of the 'physical'

My suggestion is to go back to the theory-based conception of physical, and examine some ways out of Hempel's dilemma. One alternative for the physicalist is to resort to an indexical definition of 'physical': *that kind of thing* physics says there is. *That kind of thing* will change and develop with the progress of physics and so will our physicalist commitments. This will create an open-ended definition for physicalism in which 'what is physical' changes and makes progress along with the progress and changes of physical sciences. Consequently, physicalism becomes a floating doctrine. Indexical physicalism becomes a family of theses, each member individuated by an indexical. But then we turn into Hempel's second horn of the dilemma: futures members of the family may render physicalism as a trivial thesis.

However, the open ended character allows the presence of disembodiment (minds without bodies) within a physicalist picture. Future physics might include properties which were once classified as non-physical properties, but are in the future classified as physical properties. Nevertheless, physicalists

can resist the inclusion of such entities, since there is no strong empirical evidence for the existence of ghosts or parapsychological phenomena, it is very unlikely that someday they will be granted a physical status, for now physicalism should ignore such possibilities.

Another response close to the suggestion above is to insist that present-day physics is indeed complete or that it is at least rational to consider it as complete. This is proposed by Lewis:

> It is a task of physics to provide an inventory of all the fundamental properties and relations that occur in the world. (...) We have no a priori guarantee of it, but we may reasonably think that present-day physics already goes a long way toward a complete and correct inventory. And we may reasonably hope that future physics can finish the job in the same distinctive style. (...) if we optimistically extrapolate the triumph of physics hitherto, we may provisionally accept that all fundamental properties and relations that actually occur are physical. This is the thesis of materialism. (Lewis, 1994: 51-2)

There is no structured argument to deny the first horn of Hempel's dilemma, rather, it is more of an intuition about the way we already treat physics: the intuition that it is rational to believe that present-physics is already complete. In fact, Lewis thinks that this is our attitude towards physical sciences, this is how we already proceed. Of course, there will be scientific progress which will lead to additions to the current physical science. However, the hypothesis is that no addition would be substantive enough to significantly change the face of physics. So it seems rational to preserve theory-based definitions. In the end, this is a pragmatic choice. True, there is an open-ended definition for 'physical', if physical is what is described by the ever changing physical sciences. But I do not see that as threatening physicalism. The response consists in taking the dilemma's first horn and denying its consequences: we grant that present physics is already a complete theory in the sense that new additions will not drastically change the theory.

2.3 The relation question

Now that we have settle on answers to the two first questions—the scope and the base question—we are getting closer to an adequate definition of physicalism. With the slogan 'everything is physical' we actually mean that *all instantiated properties* bear some ontologically important relation with physical properties, understood, roughly, as the properties determine by the language of the physical sciences. Now we ask what is the relation between instantiated properties in general and instantiated physical properties in particular. To respond to the relation question, we need to find the core commitments of physicalism, a *minimal physicalism* from which all versions of physicalism proceed. The dominant view among philosophers of mind is that *psychophysical supervenience* captures the most basic sense in which everything is claimed to be physical: everything is physical if and only if all properties supervene on physical properties.

2.3.1 Supervenience Physicalism

Supervenience is a dependency relation between low-level properties and high-level properties. To have an intuitive grasp of this relation, it is worth to look at how supervenience relations obtain beyond the mind and body interaction. Let us think of the global properties of a picture and the pixels that compose the picture. A picture that shows, say, the aurora borealis is composed of pixels, small dots arranged in a certain manner, so that when we stare at it from a certain distance, we see the aurora borealis. The image we see–many colors spreading through the sky–is the global property (high-level properties) of the picture, whereas the pixels are its base properties (low-level properties). Any change in the global properties (image) of the picture requires a change within the pixels of the picture, and not the other way around. The global properties supervene upon the pixels on the picture and they stand in an asymmetric relation: the former depends on the latter, but the latter does not depend on the former. In its slogan form: *there cannot be an A-difference without a B-difference.* Where A stands for the supervenient properties and B for the base-level properties. A copy of a painting will be identical to the original painting only if its lower-level properties are identical

[3]. If I am able to reproduce stroke by stroke, molecule by molecule one of Kandinsky's Compositions, that picture will be identical to the original. It is sometimes said that aesthetic properties are also supervenient properties. The arrangement of the dark and clear spots on the canvas is what makes the painting beautiful. The same relation is ascribed by moral naturalists to moral properties. Indeed, the notion of supervenience was first introduced in the context of the metaethical debate to explain a sort of normative naturalism, which argues that the normative properties supervene on the natural properties.

Now, in the philosophy of mind context, it is also claimed that physical properties and mental properties stand in a supervenience relation: the global properties (high-level) are mental properties that supervene on physical (low-level) properties. Supervenience physicalism is 'the claim that if you duplicate our world in all physical respects and stop right there, you duplicate it in all respects.' (Jackson 1998: 12) Following Jackson's formulation of supervenience physicalism (1998):

(1) Any world which is a *minimal* physical duplicate of *our* world is a duplicate *simpliciter* of our world.

The restriction of the supervenience thesis to *our* world (i.e. the actual world) is required because physicalism is a contingent thesis. Our world is physically determined, but things might have turned out differently: cartesian worlds (worlds with non-physical properties-ghosts, spirits etc.) are not impossible. So the claim is that, at least in our world, physicalism is true: given the restriction to actuality, a physical duplicate of the *actual world* is necessarily a duplicate simpliciter of our world. This idea is also formulated by Lewis (2004: 88): 'But we materialists usually think that materialism is a contingent truth. We grant that there are spooky possible worlds where materialism is false, but we insist that our actual world isn't one of them.' Once we restrict supervenience physicalism to actuality, we can say that the

[3]'Identical' in use here is in the sense of indiscernible instead of numerically different.

physical metaphysically necessitates the mental.[4]

There are reasons to believe that (1) is the proper formulation of minimal physicalism, that is, all kinds of physicalism are committed to (1). Supervenience physicalism defines the most basic physicalist position. To see this, we shall compare supervenience physicalism with two other positions which may be taken to be expressions of physicalism: token and type identity theories. Later we will consider some objections to supervenience physicalism. For now, however, what we want to ask is whether we can capture the intuitive idea of physicalism (that everything is physical) in terms of either of these two alternative theories. At this moment it is important to have in mind that supervenience physicalism, as the minimal requirement of any physicalist theory, is somewhat neutral regarding the mind-body theory in use, meaning that it is compatible with a couple of incompatible theories such as identity theory, emergentism, eliminativism etc..

2.3.2 Identity theories

According to the identity theory, our mental states are identical to our brain states, so believing that Lübeck is in Germany, that apples are red, or desiring apples are all states of brains. Psychophysical identification is inspired by scientific identifications such as the discovery that water is H_2O. The idea is that there is one phenomenon described in two different ways. The identification is established in virtue of the transitivity of identity between the phenomenon in question, its causal role, and the occupant of that role: heat is whatever occupies a certain causal role R; molecular motion is the occupant of causal role R; so, by transitivity of identity, heat is molecular motion. *Mutatis mutandis* for mental states and brain states: pain is the occupant of the causal role R, the occupant of causal role R is brain state B, so pain is the brain state B.

---

[4]The minimality requirement is introduced to prevent the duplicate of non-physical events such as miracles. It is not enough to consider a physical duplicate of the actual world because we risk duplicating a world physically like our phenomenally different. We want to duplicate only minimal physical aspects of the world.

Type-token distinction There are two kinds of psychophysical identification based on two ways of classifying individual things: consider a book and its copies. We can say we have read the same book by Thomas Mann, 'Death in Venice', although we have read different copies of the same title. 'Death in Venice' is the book-type and its copies are the tokens of the book-type. Or consider the question: How many letters are in the word 'apple'? We can count the *tokens* of *types* of letters contained in the word: a, p, p, l, e (five letter-tokens) or we can count the *types* of letters: a, p, l, e (four letter-types). Tokens are occurrences of a certain type. In psychophysical identifications we may identify (i) states with tokens of physical states or (ii) types of such states. This distinction applied to mental states yields two ways in which states can be conceived: one can follow Davidson and conceive of states as concrete *events* (particulars/occurrences) [5]. Considering mental causes as events will generate token physicalism whilst considering mental causes as properties will result in type physicalism. Let us consider:

> *Token identity theory*:
> For every actual particular (object, event or process) x, there is some physical particular y such that x = y.

The identification in token theory is between events (actual particulars) rather than properties. The main issue with token identity theory is that it is consistent with property dualism, thus not strong enough to be a proper physicalist thesis. Token identity theory makes a claim about actual items only: it establishes no modal relation between the mental and the physical. So the whole truth is in the scope of an actuality operator (not just particulars). The token identity theory allows the possibility of a duplicate of our world with no mental or phenomenal properties instantiated, so token identity thesis could be true even when supervenience fails. The fact that two particular events actually have distinct mental properties does not rule out

[5]Events are roughly things that happen, or an occurrence of a process, such as births and deaths, thunder and lightening etc.

their physical identity in close-by possible worlds. But supervenience does rule out such close-by possibilities. Thus token physicalism does not entail supervenience. That token physicalism does not entail supervenience and that it is consistent with property dualism makes it an unsuitable candidate for expressing a physicalist theory.

Supervenience physicalism also does not entail token identity theory. Token identity theory claims that for every particular, there is some physical particular to which it is identical. And we have already seen certain problems that arise when we take the domain of physicalism to be particulars instead of properties. According to the token identity theory, there must be a particular physical object to which, say, a complex object like the Goethe University is identical. But it is very difficult to say what particular physical object is identical to the Goethe University, perhaps there is none. Supervenience, by itself, does not impose this sort of reductive requirement, rather it only claims that the university is dependent on or determined by physical properties. So, supervenience physicalism does not imply token physicalism.

The type identity theory, on the other hand, refers to identity not between events or processes (considered as particulars) but between types of events of processes.

> *Type identity theory*:
> For every actually instantiated mental property F, there is some physical property G such that F=G. (Stoljar SEP)[6]

This formulation is evidently not consistent with property dualism. Then, contrary to the token identity theory, it implies the supervenience thesis: if every property instantiated in the actual world is identical with a physical property, then a world physically identical to our world will be identical to it *simpliciter*. Type physicalism entails supervenience but not the other way around. For supervenience is a contingent thesis; so it is consistent

[6] 'Actually' indicates that the type identity theory, like physicalism in general, is meant to be a contingent thesis about our world.

with the (far-away) possibility of disembodied mental properties (Cartesian worlds), whereas type identity physicalism is not. Presumably, the world could have turned out differently such that there could be disembodied souls wondering around our universe. But the sort of psychophysical identity involved in type identity theories is of the necessary kind. For this reason, type identity theory is inconsistent with the possibility of disembodiment. Hence supervenience does not entail identity theory.

Multiple realizability objection One problem with the type identity theory is that it does not cover cases in which very different physical states may occupy the same causal role characteristic for a certain mental state like pain. Consider pain in a horse. It is plausible (or so we may assume for the sake of argument) that the occupant of the pain-role in a horse is different from ours, given the significant difference between our organisms. Let us call the occupant of the causal role of pain in horses 'stimulation of d-fibers' whereas the occupant of the causal role of pain in humans is stimulation of c-fibers. If pain is stimulation of c-fibers and also stimulation of d-fibers, then (by reflexivity and transitivity) stimulation of c-fibers is stimulation of d-fibers, and that is false. Different types of state might occupy the pain-role in different organisms. Type identity cannot allow for that role to be multiply realized.[7]. Against this objection the type identity theorist could turn to the token identity theory, but we have seen that this version of identity theory does not yield an acceptable physicalist position. A solution to the multiple realizability objection is not to reject physicalism altogether, but rather to reject the identity theory.[8] We then obtain theories that are

[7]There is a better way to respond to the multiple realizability objection from the perspective of a type identity theory. One may want to finely-grained the *relata* in the identity relation. We should thus identify pain in humans with stimulation of c-fibers and pain-in-horses with stimulation of d-fibers instead of plain pain, and so on for other cases.

[8]The most prominent argument against type-identity theory is Kripke's (1972) conceivability argument, which will be addressed in detail in the next chapter when we discuss anti-physicalist arguments.

subsumed under the heading of 'non-reductive physicalism' like functionalist and emergentist theories. The functionalist approach individuates mental phenomena according to their causal roles.

Nonetheless, supervenience is admittedly a weak thesis. Kim (2000) goes even further and says that besides being weak, it does not provide a satisfactory account of the mind-body problem. It merely states a pattern of property covariation between the mental and the physical and points to the existence of a dependence relation, being silent on matters like the nature of that relation, that it is fails in explaining what the relation is. What favors for this deflationist account of supervenience is that supervenience itself seems to be a commitment of different and conflicting physicalist positions. Type identity theory implies supervenience, as well as realization physicalism—the view that the mental is physically realized—and epiphenomenalism—if two individuals differ in some mental respect, they have to differ in some physical respect—among other theories are all consistent with psychophysical supervenience. So, the supervenience thesis endorses views that have physical processes at the bottom of mental processes and rule out views that allow the mental world to float freely, unconstrained by the physical domain' (Kim 2000: 15). This certainly is a core commitment of physicalism. The supervenience thesis looks like the right candidate for *minimal physicalism*. Surely one can strengthen supervenience to obtain stronger physicalist theories. But the very neutrality of supervenience is what qualifies it as the key ingredient in a minimal physicalist answer the relation question.

We have now provided responses to the three branches of the question of how to interpret the physicalist slogan: *everything is physical*. The answer to the scope question was to restrict the domain of the quantifier to instantiated properties, whereas the response to the relation question was to formulate the dependence relation of supervenience, which claims that *any world which is a minimal physical duplicate of* our *world is a duplicate* simpliciter *of our world*. At last, in response to the base question of what is a physical property, we have concluded that although the theory view presents some important shortcomings, it is still the most promising way to define the domain of the physical. Hence, the answer to the interpretation question the physicalism is: every instantiated property supervenes on properties

expressed by a true physical theory. This definition entails that physical properties metaphysically necessitate all properties simpliciter.

3 The truth question: why is physicalism so plausible?

3.1 The causal argument

The canonical argument for physicalism is the so-called *causal argument* or *argument from causal closure*. It is based on two independently plausible ideas. First, mental events have causal powers. My wanting to raise my arm, causes the raising of my arm. Second, the domain of physical events is (backwardly) causally closed: any physical event has a complete purely physical cause. If we combine these two ideas, we find that a physical event like raising my arm is causally overdetermined, in having, first, a mental cause, and, second a complete physical cause. Note that the physical cause is assumed to be complete so that it does not need to collaborate with the mental cause to produce the effect. Now causal overdetermination may sometimes happen, as in the famous example of two bullets shattering simultaneously a bottle. But such cases—if they are cases of genuine overdetermination—are rare. What makes it unattractive to combine the two ideas is that it makes *every* case of mental causal a case of overdetermination. That is to say, the combination of the two ideas is incompatible with a third idea, the causal exclusion principle, that one complete cause of an event excludes all other, distinct causes—perhaps notwithstanding rare cases of overdetermination. The only way to hold the first and the second idea consistently with the causal exclusion principle it to deny that in what we describe as mental causation of a physical event, the mental cause is distinct from the physical cause of that event. Mental causes are physical causes. We can briefly summarize the causal argument:

(P1) All physical events have a complete physical cause.

(P2) There are mental events c and physical events e such that c causes e. (Call such events physically efficacious.)

(P3) For all events c and e, if c is a complete cause of e, then all causes c' of e are part of c.

(C) Mentally efficacious events are parts of physical causes and hence, themselves physical.

Premises (P1) and (P2) tells us that certain effects have a mental cause *and* a physical cause. (P3) tells us that they do not have distinct causes. This leads to the conclusion that mental events are physical events. (P1) is also called the completeness of physics (Papineau 2002). The fact that physics is complete should not be confused with the idea that physics can explain everything and the idea that physics is a complete science. The label 'complete' only means that physical causes are sufficient and complete to bring about all physical effects. Serious problems arise when tried to deny the completeness of physics. One may hold a dualist interactionist view, according to which behavior has a two-way causation. Our mental states produce physical effects and physical causes produce mental effects. However, this view violates (P3) the causal closure premise.

Alternatively, one can argue against the causal argument by denying the causal efficacy of mental states thereby embracing an epiphenomenalist position: the view that mental events are caused by physical events but are causally powerless. However, this is inconsistent with the causal efficacy premise (P2). The causal efficacy premise is very intuitive. It seems evident that I go to the cafeteria to buy a cup of coffee because I feel sleepy and I believe that some caffeine will do the job of keeping me awake, I raise my arm because I decide to do so. This is indeed a very strong intuition and to deny it seems at least extravagant. If mental efficacy is an illusion, it is a very powerful illusion. For example, human agency, as we understand it, requires mental causation. I change my behavior based on the acquirement of new beliefs and discarding of old beliefs. So, mental causation is at the centre of our folk psychology. Hence, adopting epiphenomenalism requires, first an error theory that explains why intuitions to the contrary have such a strong hold on us, and, second, a reform program, replacing the mistaken folk psychology be the truth of the matter in such a way that, say changing beliefs in the interest of acting successfully still is a rational enterprise.

Furthermore, it is sometimes said that epiphenomenalism is an empirically implausible position inasmuch as effects without causal powers is not a phenomenon one might find in nature. One might say that science provides no example of 'causal danglers'. Besides that, even if epiphenomenalism and physicalism were to share the same explanatory force, our ordinary scientific methodology would advise us to choose the most simple view, rather than the ontologically profligate story which has the conscious states dangling impotent from the brain states.

A third possibility for the dualist is to deny (P3) (i.e. the exclusion principle) and embrace overdetermination. As explained above, overdetermination is the phenomenon that there can be two or more independent causes of an effect, each sufficient by itself for bringing about the effect (Jackson and Braddon-Mitchell 1996: 16). A paradigmatic case for overdetermination is that in which Jones is killed by two different and simultaneously shots. Each shot is sufficient to cause the death of Jones. Although we can think of overdetermining causes in non-mental cases, like in Jones case, it is difficult to accept it as a requirement in mental cases.

Causes a and b in overdetermining cases must be sufficient to bring about effect e. However, it is difficult to entertain the analog mental case. How could I raise my arm without wanting to raise my arm (ruling out abnormal cases of involuntary movements)? Consider that cause a is my wanting to raise my arm and cause b is some neural process (to which I have no conscious). According to overdetermination, a and b are sufficient to bring about effect e (the raising of my arm). But it is hard to see how b alone could be sufficient to bring about the raising of my arm. This is just to point the disanalogy between mental cases of overdetermination and physical cases. In Jones' case, it is easy to imagine a scenario which, not a, but b is a sufficient cause of Jones' death. But in the mental scenario, it is difficult to see how excluding the mental cause a and preserving the physical cause b is sufficient to cause e.

Besides the intuitive problem, a shortcoming for overdetermination is the same problem with epiphenomenalism: we are not used to find many cases of overdetermination in nature, this should raise some suspicious, if overdetermination is, at best, a rare phenomenon, so why should we judge that

every case of mental causation is a case of overdetermination? This seems highly implausible. We are better off preserving the exclusion principle.

The problem that the causal argument presents for dualism is how to reconcile the three principles mentioned previously: (i) completeness of physics or physical closure; (ii) mental causal efficacy and (iii) the exclusion principle. The dualist view is ruled out by these assumptions, hence the only ontology that is consistent with all of them is a kind of physicalism. In fact, the causal argument seems to derive one to an identity theory: that mental causes are identical to physical causes. However, the causal argument is also an argument for other formulations of physicalism. Including supervenience argument.

* * *

The result of this chapter was to come up with a preliminary definition of physicalism more strict than that provided by the slogan 'everything is physical'. The answer to the interpretation question the physicalism is: every instantiated property supervenes on properties expressed by a true physical theory. This definition entails that physical properties metaphysically necessitate all properties simpliciter, hence P$\rightarrow$Q, where P is the complete physical description of our world and Q is a phenomenal truth. This implication is the core thesis of physicalism and we must keep that in mind in order to understand the anti-physicalist arguments, which are the topic of the next chapter.

Chapter 2

Challenge from Conscious Experience

1 Introduction

At the end of the previous chapter, we have briefly examined the positive side of the truth question: arguments that deliver plausibility to physicalism and that are widely accepted in the debates in Philosophy of Mind. In the present chapter, we will focus on the negative side of the truth question, viz. arguments against physicalism, then examine some arguments that emphasize the subjective character of consciousness as its core feature. We will now examine how anti-physicalist arguments give rise to the tension between those aspects of our everyday life (with focus on phenomenality) and the thesis of physicalism. The debate over the subjective character of consciousness, or as it is sometimes called: 'the hard problem of consciousness' (Chalmers, 1996), is considered to be the greatest challenge to physicalism. Many philosophers posit that this is a matter that cannot be solved, regardless of scientific progress, for it is beyond the scope of what science can find out about the world. If they are correct, the consequence is that the idea of physicalism itself fails.

The term 'qualia' (singular 'quale') is used to refer to the phenomenal aspects of our mental lives, that is, what it is like to undergo some experience, such as what it is like to see red, to smell freshly ground coffee, new

mown grass etc.. However, which (sort of) mental states have qualia? It is still controversial to ascribe this peculiarity of experience to mental states such as beliefs and desires[1]. Cognitive states are paradigmatic examples of mental states with no *phenomenal feel*. My belief that 2+2=4 or that the Earth is oval has nothing that *it is like*—at least not in the same sense as feelings of pain and other perceptual experiences do. So we need to distinguish between mental states that have phenomenal feel from mental states that do not have phenomenal feel. Typically, we say that perceptual states have phenomenal feel, whereas cognitive states do not.

Physicalists need to address powerful arguments involving these phenomenal aspects (qualia) of our mental lives. These arguments are about our epistemic, subjective situation confronted with an objective view of the world. This confrontation opens an epistemic gap between the way we introspect aspects of our experience and the physical explanation of reality. We will focus on two central arguments against physicalism, namely, *the conceivability argument* (Chalmers 1996) (two versions of the conceivability argument will be addressed) and *the knowledge argument* (Jackson 1982). Both arguments consist of the assertion of an epistemic gap between the physical and the phenomenal domain, and then proceed to claiming a metaphysical gap. The conclusion is the falsehood of physicalism. These arguments turn into a contradiction the apparent difficulty of accommodating qualia within a physicalist framework.

---

[1]There is a growing debate in Philosophy of Mind as to whether phenomenality can be ascribed to intentional states such as beliefs and desires. The debate is known as both 'cognitive phenomenology' and 'Phenomenal Intentionality'. I will not engage in this matter here. On the contrary, I will consider the traditional view that ascribes phenomenality to a specific kind of state.

2 Conceivability Arguments

2.1 Descartes and Chalmers

The *locus classicus* for the conceivability argument is Descartes' Meditation VI. Descartes aims at deriving the distinctness of mind and body from an epistemic assumption; one can form a *clear* and *distinct idea* of one's mind existing without one's body. The fact that mind and body may be conceivable as distinct entities is a tool to infer that they could indeed be distinct, and the possibility of them being distinct is then used to conclude that they are indeed distinct. Therefore, physicalism is false.

(P1) I can conceive clearly and distinctively that I, a thinking thing, can exist without my extended (i.e., physical) body existing.

(P2) Anything that I can conceive of clearly and distinctly is logically possible.

(P3) If it is logically possible that X (mind) exists without Y (body), then X (mind) is not identical to Y (body).

(C) Therefore, I, a thinking thing, am not identical with my extended body.

The problem with this seminal formulation is the vagueness of the notion of a 'clear and distinct idea' (conceivability). If one conceives clearly and distinctly X without Y, then it is possible that X exists without Y for God must be able to produce any distinction one is able to conceive in one's mind. The link between conceivability and possibility depends on the proof of God and God's ability to produce anything that is conceivable. However, Descartes does not present a detailed constraint to the type of conceivability that leads to metaphysical possibility. Thus it is not clear which kind of conceivability would lead to possibility. This puts into question the core assumption that conceivability leads to metaphysical possibility.

Faced with that difficulty, Chalmers formulates an updated version of the argument with the aid of his interpretation of the two-dimensional semantics in order to build a solid link between conceivability and possibility.

2.2 Two-dimensional argument against physicalism or the Zombie argument

Let us consider a creature—physically and functionally—identical to me but with no subjectivity. The creature behaves just like me; it is molecule for molecule identical to me, but it lacks the subjective character of consciousness; it is experientially empty; it has no feeling such as 'Oh! So, this is what it is like to see the beach for the first time'. While I have a certain feeling when I taste Swiss dark chocolate, my zombie-twin would react just like me: she would present the same behavioral responses, but would lack the special feeling I get when I taste Swiss dark chocolate. The creature processes the same information as I do: she processes the same perceptual data as me, thus she produces the same behavioral outputs as I do. Nevertheless, she lacks the phenomenal experience which I possess. As Chalmers (1996: 95) puts it: 'there is nothing like to be a zombie.'

It is plausible to deny the *nomological* possibility of zombies, since there are not and there could not be such creatures in the actual world. Nonetheless, we are dealing here with the *metaphysical* possibility of a creature with absent qualia. The idea that physicalism is incompatible with the metaphysical possibility of zombies is pretty straightforward: If I had a zombie-twin (a physical duplicate of me) then there could be, according to physicalism, no difference *simpliciter* between us. Contrapositively: If there were a difference at the phenomenal level, then, according to physicalism, we could not be perfect twins. The conclusion is that these phenomenal states are not entirely physically determined. The metaphysical possibility of an absent qualia creature immediately violates our formulation of minimal physicalism:

(1) Any world which is a *minimal* physical duplicate of *our* world is a duplicate *simpliciter* of our world.

To consider the possibility of zombies is to consider that (1) is contingent. Although physicalism is a contingent thesis, (there are possible worlds in which physicalism is false) the restriction to actuality must yield a necessary condition: in *our* world, once every physical respect P is settled,

every mental respect Q will also be settled, rendering us the psychophysical conditional: If P then Q. Considering the metaphysical possibility of the existence of zombies is to consider a duplicate of our world lacking some special feature. It is to consider that the conditional is false, and that alone violates (1). Following Chalmers (1996) we can now see how the zombie argument is formulated:

(P1) P&¬Q is conceivable.

(P2) If P&¬Q is conceivable, P&¬Q is metaphysically possible.

(P3) If P&¬Q is metaphysically possible, physicalism is false.

(C) Physicalism is false.

Let P stand for all physical truths in the world and Q stand for all the phenomenal truths. In (P1) the physical properties are kept constant, whereas the phenomenal properties vary. As explained previously, to conceive of a physical duplicate lacking phenomenal states is to conceive minimal physicalism as false. Further, as conceivability implies metaphysical possibility, the metaphysical possibility of zombies is inconsistent with physicalism. The conceivability argument is clearly valid. Physicalists need to show that at least one of the premises is false. Objections to the argument will typically question the first two premises: (i) are zombies conceivable? If they are conceivable, (ii) does it follow that they are possible? The proponent of the conceivability argument must answer positively to both questions. Consequently, physicalists will say no to either the first or the second question. We shall now examine the possible physicalist reactions.

2.2.1 Zombies are not conceivable

One possibility is to reject (P1) which asserts the conceivability of zombies. Analytic functionalists choose this path by claiming that the meaning of phenomenal terms can be analyzed in functional terms. Terms that designate phenomenal states, such as 'pain', have their meaning fixed by whatever plays the functional role of pain: whatever causes behavior typically related

to pain. If what plays the functional role of pain is the physical stuff underneath it, then the physical stuff, for example, 'stimulation of c-fibers', plays that functional role in any minimal physical duplicate of our world. If this is the case, there would be no way of conceiving zombies; the relevant functional role of 'Q' is played by 'P'; so, whatever instantiates P will also instantiate Q. Therefore, the conceivability of zombies is rendered impossible. Nevertheless, analytic functionalists must explain at least why zombies *seem* conceivable.

Those physicalists might suppose that our conceiving of zombies is somewhat deficient, not meeting ideal standards. The notion of conceivability, which the proponent of the argument employs here, is the one that abstracts from our limited cognitive capacities. Since conceivability at stake here is not ideal conceivability, we are subject to error regarding the conceivability of many propositions. One paradigmatic example is the conceivability of complex mathematical truths. It is said of the Goldbach conjecture that its truth *and* falsity are conceivable, but that cannot be the case since one of the two options is a priori truth and the other a priori false. These only seem conceivable because our limited reflection skills make that error. The problem with this kind of response is that it presupposes the complicated notion of idealized conceivability. This problem will also be present when we discuss different dimensions of conceivability further in this chapter.

To reject (P1) one must claim Q is a priori deducible from P by an ideal reasoner. I will agree with Chalmers regarding the strong intuition behind the truth of (P1), and against the a priori deducibility of Q from P. It is quite intuitive that given the complete microphysical conception of the actual world, the falsity of 'Q would not be ruled out, despite the level of knowledge and cognitive capacity the reasoner has.

2.2.2 Link between conceivability and possibility

The most popular physicalist response to the conceivability argument consists in rejecting the link between conceivability and possibility. There is, however, some initial plausibility in linking conceivability and possibility. There are different kinds of possibilities, there is metaphysical possibility,

logical possibility, nomological possibility etc.. The kind of possibility that is relevant to us is metaphysical possibility. Consider, first, the contrast between nomological and metaphysical possibility. It is conceivable that tele-transportation exists but the existence of tele-transportation is surely not *nomologically* possible. The laws of physics that rule our world do not allow for such possibility. However, if the world were ruled by different physical laws, tele-transportation would be a tenable possibility. This is a metaphysical possibility: If the world were to be different in such-and-such ways, x would be the case, so x is metaphysically possible. If there were no way the world could have turned out that x would be the case, we say that x is not metaphysically possible. Examples of metaphysical impossibilities are contradictory thoughts: there is no way a world could turn out in which square circles would be metaphysically possible.

However, many philosophers reject this link mainly by focusing on counterexamples to the alleged bridge between conceivability and possibility. Two famous examples are: (i) the Goldbach conjecture or its negation are conceivable but one of them is impossible and, (ii) a posteriori identities are conceivable though metaphysically impossible. The issue with a posteriori identity statements is that, after taking Kripke's considerations into account regarding the identity between rigid designators, we are inclined to agree that every identity is necessary, therefore, this makes them metaphysically impossible to separate. Nonetheless, it is quite reasonable to imagine them coming apart, i.e. to imagine that Hesperus is not Venus or that Gold is not Au79. These counterexamples suggest that we are dealing with a very specific notion of conceivability that needs to abstract away from any rational limitations that constraint our reasoner's abilities to conceive, among other things. This is the point where Chalmers' version of the conceivability argument differs from that of Descartes. In order to argue in defense of the conceivability-possibility link, one needs to specify what notion of conceivability yields metaphysical possibility. Chalmers (2002) makes an inventory of kinds of conceivability and concludes that only one kind of conceivability is safe guide to metaphysical possibility. This is where Chalmers' two-dimensional framework comes in to explain what is required to safely pass from conceivability to possibility.

2.3 Two-dimensional semantic framework

'Two-dimensional semantics' designates a set of semantic theories within intensional semantics. The latter may be characterized by the five core theses: (i) the meaning of linguistic expressions is essentially representational; (ii) the representational content is identical to the truth conditions of the sentence; (iii) truth conditional contents amounts to the distribution of truth values in possible worlds. In each possible world, the sentence will return a truth value; (iv) extensions are compositional; (v) intensions are also compositional. Thus, meanings are intensions. Two-dimensional semantics postulates in addition that the truth-value of certain sentences holds a double dependence vis-à-vis possible worlds. To represent the two ways that truth-values depend on possible states of affairs, two-dimensional semantics systematically assigns a pair of intensions to each linguistic expression: a *primary intension* and a *secondary intension*. This is a general commitment of the framework, notwithstanding the fact that there are many interpretations for the formal apparatus in question.

Let us focus for now on the general formal apparatus of two-dimensional semantics.[2]. First, one-dimensional semantics ascribes only one intension to each linguistic expression. This intension picks out the way truth-values depend on facts, whereas, the two-dimensional framework aims to pick out another kind of dependence, viz. the truth value of a sentence vis-à-vis what it means or conveys. The two-dimensional semantics generalizes the double-indexing strategy developed to deal with certain expressions, such as indexicals, demonstratives and tense terms ('here', 'now', 'that'). David

[2]There is a variety of two-dimensional semantic theories that diverge in specific points. Nevertheless, they share the fundamental idea: the double-dependence of truth-value in face of states of affairs. Stalnaker, for example, takes the framework as a model for meta-semantical facts about language thus endorsing a pragmatic version of two-dimensional semantics. Chalmers, on the other hand, views the framework as representing semantic facts. The latter seeks to establish a priori connections between modality and meaning. We will focus on the so-called rationalist or epistemic version of two-dimensionalism. Exponents of this version are David Chalmers and Frank Jackson. This version seeks to vindicate the conceivability argument against physicalism.

Kaplan's semantic framework (1989) is widely used to explain conventional semantic rules governing context-dependent expressions such as indexicals and demonstrative terms. These differ in content depending on the contextual usage of the expression. Logicians working in tense and modal logic use two-dimensional semantics to characterize the logical properties of operators such as 'now', 'actually', and 'necessarily'. Such applications of the two-dimensional framework are uncontroversial.

The more ambitious interpretations of two-dimensional semantics generalize the uncontroversial applications to apply to *all* sort of expressions. Such ambitious interpretations seek to isolate a priori aspects of meaning. This generalization departs from two core ideas: there are two ways a linguistic expression depends on possible worlds. First, the primary extension of an expression depends on the nature of the actual world in which the expression is uttered. Second, the secondary extension of an expression depends on the nature of the world in which the expression is counterfactually evaluated. Corresponding to these two kinds of dependence, there are two kinds of intension. An intension is a function that takes an expression relative to a world. So the intension of the expression e is a function from a possible world w to the object that satisfies e. The extension of a sentence is a truth-value in a particular world, whereas the intension is the proposition expressed. The two-dimensional framework stipulated two kinds of dependences of expressions on possible worlds, we have two ways of considering possibilities: (i) the possibilities represent the way the actual world might have turned out to be, which is equivalent to 'considering a possibility as actual' or to consider that the world we are evaluating is *our* world. (ii) The other way to consider possibilities is to 'consider a world as counterfactual'. In the latter, the actual world is already fixed and the extension of the linguistic expressions will have the same truth value at counterfactual worlds as they do at the actual world. Or to put it in Chalmers' vocabulary, the possibility that represents the world as actual is a primary possibility and the possibility that represents the world as counterfactual is a secondary possibility.

Considering a possible world w as actual is to consider the possibility that w is *our* world, that is, to consider the possibility that the actual world

could have turned out to be different. Thus, it is possible to consider that our world is such that the watery stuff in it is XYZ and not H_2O. This is how Putnam's doppelgänger at Twin-earth would perceive his world when reflecting on the reference of Twin-water across worlds. It is a reflective exercise that takes our actual world and raises the hypothesis that other worlds could also be actual. Then we should evaluate meanings across w as if w were actual. By contrast, considering a world as counterfactual is to think of a different possibility under the condition that the meaning of the expression is fixed in the actual world. Therefore, according to the two-dimensional semantics, if an expression is evaluated relatively to a world w, it has two intensions as a result, depending on how w is conceived (actually or counterfactually). As Chalmers puts it, it is primarily possible that water is XYZ but it is not secondarily possible that water is XYZ, but it is secondarily possible that water is H_2O.

The two kinds of intensions characterized above represent two dimensions of meaning. At this point, it is convenient to mobilize semantic matrices to visualize the double-dependence that two-dimensional framework stipulates and each of the corresponding intensions. Consider the following sentence containing an indexical term:

(1) I am sick.

The relevant possibilities are: at the world i Mary is the utterer and at the world j Peter is the utterer of (1). Mary is sick in both worlds, i and j: Peter is not sick in neither i nor j. We can represent the double-dependence of the truth value of (1) by mobilizing the semantic matrix:

	i	j
i	1	1
j	0	0

The worlds represented in the vertical axes are worlds considered as actual, or primary possibilities. The worlds represented in the horizontal lines are

considered as counterfactual. The secondary intensions of (1) are represented in the horizontal lines of the matrix: If we consider i as actual, then (1) is true are j considered as counterfactual. At the world i, 'I' identifies Mary, so the term 'I' rigidly designates Mary in j. Mary is sick in both i and j, since the actual world is where the reference of the terms is fixed, once the reference is fixed, we are in a position to evaluate the sentence in a counterfactual world. Nevertheless, if we consider j as actual (the second line of the matrix), then (1) is false at i considered as counterfactual. At j, the term 'I' picks out Peter, therefore, the term 'I' rigidly designates Peter at i. The primary intension of (1) is represented by the diagonal of the matrix. The primary intension is true at i, since, by considering i as actual and by evaluating (1) in the actual world, the output will be the 'True'. In the actual world, 'I' picks out Mary and she is sick at 'i'. When considering j as actual, 'I' picks out Peter in j and Peter is not sick in j, hence the result is 'false'.

Possible worlds w play the role of contexts of utterances that determine the extensions in w. The extension of an indicative sentence is a truth value: true or false. The primary extension of a sentence depends on the world in which the sentence is uttered. In this case, the context of utterance determines the truth-value of the sentence. The secondary extension of an expression in w depends on the worlds considered as counterfactual — it is no longer the nature of the world in which the sentence is uttered.

Now consider the sentence (2):

(2) There is water here.

The matrix below represents the two dimensions of meanings of the sentence (2). The possibilities w,v,u represented in the vertical axes are *worlds considered as actual*, and the possibilities represented by horizontal lines are *worlds considered as counterfactuals*. The relevant possibilities are: at w 'water' designates H_2O, at v 'water' designates XYZ and at u 'water' designates KLM.

	w	v	u
w	1	0	0
v	0	1	0
u	0	0	1

The matrix represents the two ways to evaluate a sentence: the secondary intension of (2) is a function from worlds considered as counterfactual to extensions. Keeping the actual meaning of 'water' (H_2O) and evaluating (2) in counterfactual worlds will yield the following results: 'water' designates H_2O in every possible world so (2) will be true in w but false in v and u considered as counterfactual since only in w 'water' designates H_2O. The secondary intension of (2) is represented in the line of the matrix.

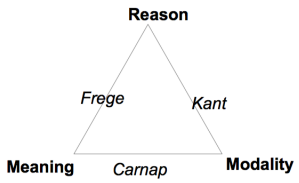
The primary intension of (2) is represented by the diagonal in the matrix. Considering w, v and u as actual, (2) will be always true, since regarding primary intension, 'water' will pick out whatever plays the watery role in the circumstances of evaluation, and at w is H_2O, at v is XYZ and at u is KLM. We can observe that the primary intension is represented by the intersection points between worlds considered as actual and worlds considered as counterfactual.

The possibilities represented in the vertical axes are considered as actual and the ones represented in the horizontal lines are counterfactual. That is depicted in the vertical axes of the matrix. Now the hypothesis could be potentially made that v (XYZ-world) is the actual world, then 'water' designates XYZ in v, so: there is water in v but not in w or u (which are worlds that do not contain XYZ). Now consider u (KLM-world) as actual, then 'water' designated KLM in u, so: there is water in u, but not in w or v. The primary intension of (2) is always true when the sentence is evaluated in the same world that is being considered as actual. Consequentially, it returns true in the diagonal of the matrix where there is an intersection between worlds evaluated and worlds considered as actual. The primary intension of the sentence picks out whatever plays the watery role regardless of the extension that the sentence has in the actual world. The secondary intension of the sentence, on the other hand, picks out the extension the

sentence has in the actual world and evaluates it in different counterfactual worlds.

This is how the general account of two-dimensional semantics works, the generalized version of the framework generalizes the application of the semantics to all linguistic expressions. Only that generalized version is relevant for our purpose since it aims to produce the link between conceivability and possibility. This is known as the *epistemic interpretation* or the *rationalist project*, which is Chalmers' specific interpretation of the framework. He proposes an interpretation of the two-dimensional semantic framework that aims to restore the metaphysical connections between modality and reason shredded by the allegedly existence of Kripke's so-called modal hybrids. To illustrate, Chalmers recalls the picture of a Golden Metaphysical Triangle between three great pillars of metaphysics: reason, meaning and modality, as an era of metaphysical harmony.

Kant connected reason and modality by the thesis that what is necessary is a priori and what is a priori is necessary. Frege then connected meaning and reason by distinguishing between sense (*Sinn*) and reference (*Bedeutung*) in order to account for the difference in cognitive significance of certain identity statements with co-referential terms. Given that Frege's notion of sense is somewhat vague, Carnap proposed an explication of Frege's sense in terms of intensions that marks the last connection between meaning and modality. Carnap suggests that two expressions have the same intension if and only if they are necessarily co-extensive. 'Renates' and 'Cordates' are co-referential expressions, but do not necessarily designate the same entity. It is at least possible that renates are not cordates. So 'renate' and 'cordate' have the same extension but different intensions. 'The result was a golden triangle of constitutive connections between meaning, reason, and modality' (Chalmers 2006: 55):



Kripke argued against the Kantian thesis by pointing out the existence of the necessary a posteriori which was ignored by the philosophical tradition that preceded him: certain necessary connections cannot be a priori known, this is the case with identities between rigid designators. Moreover, if two expressions are necessarily co-referential then, according to Carnap's thesis, they have the same intension. But this destroys the Fregean connection between meaning and reason.

The two-dimensional semantics aims at rehabilitating the connections between reason and modality, on the one hand, and reason and meaning, on the other hand. The generalized interpretation of the framework is required to make for the appropriate link between conceivability and possibility. Vindicating the golden triangle constitutes the rationalist interpretation of the framework that is able to account for the connection between conceivability and possibility required by the conceivability argument against physicalism. Another way to express the vindication of the golden triangle is by endorsing a thesis that gives substance to the rationalist interpretation of the two-dimensional semantics, viz.:

> *Core thesis*:
> For any sentence S, S is a priori if and only if S has a necessary primary intension.

As primary intension is captured by the diagonal propositional, any propositional concept that has a necessary primary intension will be a priori. Included in the central thesis are all the relevant elements of the reconnected triangle: modality, meaning and reason. To illustrate the core thesis, some additional matrixes for a priori sentences will be elaborated. Consider:

(3) Equilateral triangles are equiangular.

	w	v	u
w	1	1	1
v	1	1	1
u	1	1	1

(3) is always true in every world considered as actual or considered as counterfactual. The primary intension and the secondary intension are necessary. Now consider the two-dimensional analysis of the following paradigmatic case of contingent a priori.

(4) The stick in Paris is one meter long.

It is stipulated that the expression 'meter' designates the length of a stick located in Paris. Hence, we can know a priori that this stick is one meter long. However, the proposition expressed by (4) is contingent. One could have stipulated that a different stick shall be called one meter.

	w	v	u
w	1	0	0
v	0	1	0
u	0	0	1

The relevant possibilities are: w is the actual world where the stick has one-meter long; v is a counterfactual worlds in which the stick is 1.1 meter and u is also a counterfactual world in which the stick is 0,9 meter. We can see in the matrix that the secondary intension (represented horizontally) is contingent and that the primary intension, represented diagonally, is necessary. (4) is a priori, for the primary intension which is necessary, and contingent for the secondary intension is contingent. The diagonal in the matrix represents the epistemic character of the sentence, which is a priori. The horizontal lines represent the modal character of the sentence, which is contingent. The two-dimensional analysis of Kripke's modal hybrid is that there is no such thing as propositional hybrids, rather, they are explained as sentences which have two propositions associated, one contingent proposition (secondary intension) and one necessary proposition (primary intension). This allows us to represent the Kripkean necessary a posteriori. Consider the paradigmatic case (5):

(5) Hesperus is Phosphorus.

	w	v	u
w	1	1	1
v	1	1	1
u	0	0	0

The relevant possibilities are: Hesperus and Phosphorus designate Venus at w, at u they both designate Mars and at v Hesperus designates Mars and Phosphorus designate Venus. As a result, we have the primary intension (diagonal) a posteriori contingent and the secondary intension (horizontal) necessary a priori (true in w and v and false in u). This represents two different propositions associated with the sentence (5). Again, a necessary a posteriori sentence does not express one necessary a posteriori proposition, rather, it expresses two propositions: a posteriori contingent proposition and a priori necessary proposition. This is the way that two-dimensional semantics seeks to repair the connection between modal and rational domains shredded by Kripke's analysis of modal hybrids. Since there is no sentence that expresses only one proposition simultaneously necessary *and* a posteriori, the previous thesis remains intact: every a priori proposition is still necessary; the connection should be restored.

I have presented so far the general framework for two-dimensional semantics and one way of interpreting them. This exposition was required since Chalmers' version of the conceivability argument is grounded in his particular interpretation of the two-dimensional semantics.

2.4 Dimensions of conceivability

With the semantic apparatus in hand, we can now comprehend the conceivability argument as formulated by David Chalmers. Chalmers' argument operates by producing ontological conclusions from epistemic premises. The real work in this argument is done by bridging the epistemic and modal domains *via* the two-dimensional framework. Chalmers makes an inventory of different notions of conceivability which will serve as candidates to entail metaphysical possibility. Those new dimensions of conceivability should also accommodate widely recognized counterexamples to the link. He dis-

tinguishes between three dimensions of conceivability: prima facie vs. ideal, negative vs. positive, primary vs. secondary. This is a requirement to assess both (P1) and (P2) of the argument. We shall begin with the first dimension:

2.4.1 Negative vs. positive conceivability

> *Negative conceivability*:
> S is conceivable if and only if S cannot be *ruled out* through a priori reasoning. (p.143)
>
> *Positive conceivability*:
> S is positively conceivable when one can coherently imagine a situation in which A is the case. (p.144)

Negative conceivability explicitly resorts to the notion of a priority and yields close connections with conceptual analysis. It is defined so as to exclude any contradictory sentences such as 'round squares', 'married bachelors' etc.. Hence, per definition, any sentence which does not contain any a priori contradiction is conceivable. Examples of negative conceivability are 'water is not H_2O', 'the moon is made out of cheese', 'pigs can fly'. Conceivability is negative as it is defined in terms of *ruling out* what cannot be excluded by purely a priori reasoning. On the other hand, positive conceiving of a sentence S consists in the ability to *imagine* a coherent situation which verifies S; it is a definition centered in the faculty of imagination. Defining conceivability in terms of 'what can be ruled out a priori' is negative conceivability.

Imaginability is the core notion in the positive dimension of conceivability. How strict should we be when defining 'coherent imagination'? With what kind of reasoner are we dealing? Is it a common subject with limited cognitive capacities? Or an ideal reasoner with ideal cognitive capacities? This questions suggests an additional distinction.

2.4.2 Prima facie vs. ideal conceivability

> *Prima facie conceivability*:
> S is prima facie conceivable (for a subject) when that subject is unable to rule out the hypothesis expressed by S by a priori reasoning on *initial considerations*. (p.143)
>
> *Ideal Conceivability*:
> S is ideally conceivable when the hypothesis expressed by S cannot be ruled out a priori even on *ideal reflection*.

Prima facie conceivability is tied to the subject's contingent cognitive limitations. The absence of ideal cognitive capacities may lead to mistakes in a priori reasoning, such as judging S to be prima facie conceivable on initial consideration and then later, upon deeper reflection, seeing that S is not really conceivable. Alternatively to prima facie conceivability, we have ideal conceivability that abstracts away from the subject's cognitive limitations and requires that the utterer of S has ideal cognitive capacities. Some sentences certainly fail to be even prima facie inconceivable, such as simple mathematical falsehoods, like '2+2=5', other sentences, such as complex mathematical truths are prima facie conceivable as false but, ideally inconceivable as false.

We can combine prima facie and ideal conceivability with negative and positive conceivability. We have already considered examples of prima facie and ideal *negative* conceivability. A sentence S is *prima facie positively conceivable* if a subject can imagine a situation that she takes to be coherent (on first reflection) verifying S. A sentence S is *ideally positive conceivable* if its coherent imaginability cannot be ruled out a priori on ideal reflection. One paradigmatic case of positive conceivability in philosophy is exemplified in Descartes' conceivability notion. He claims to have 'clear and distinct ideas', which is equivalent to imagine a scenario that is coherent on sustained reflection and which verifies some sentence S.

It is clear that ideal conceivability is a better candidate to entail metaphysical possibility than prima facie conceivability. Prima facie conceivability is susceptible to failures in reasoning resulting in falsely considering a

sentence as conceivable or in failing in see its inconceivability. Examples that illustrate the asymmetrical advantage of ideal over prima facie conceivability involves the conceivability of complex mathematical truths. The inferiority of prima facie conceivability is made clear by obvious counterexamples against the entailment between conceivability and possibility. Here is one:

- The Goldbach conjecture is claimed to be both conceivable as true *and* false, but that cannot be the case since one of the two options is a priori true and the other a priori false. Both options seem conceivable due to our limited reflection skills. The distinction between prima facie conceivability and ideal conceivability accommodate this counterexample: the Goldbach Conjecture is prima facie conceivable as false *and* as true, but it is ideally conceivable either as false *or* as true.

2.4.3 Primary vs. secondary conceivability

Another counterexample to the link between conceivability and possibility, which is not accommodated by distinguishing between prima facie and ideal conceivability concerns the so-called Kripke's modal hybrids. This counterexample requires an additional dimension of conceivability:

- Kripke's notorious analysis of the necessary a posteriori yields the following results: It is claimed that necessary a posteriori sentences are conceivable as false despite their being metaphysically impossible: 'Water is not H_2O' is conceivable as true, but, because the terms involved in the identity statement are rigid designators, the statement is necessarily false, hence not metaphysically possible.

According to Chalmers, there is a sense of conceivability in which the sentence 'water is not H_2O' is conceivable, and a different sense in which it is not conceivable. This is where the two-dimensional framework comes in, to aid in the distinction between an additional dimension of conceivability: primary and secondary conceivability.

Primary conceivability:
S is primarily conceivable when it is conceivable that S is actually the case.

Secondary conceivability:
S is secondarily conceivable when S conceivably might have been the case.

There are two senses of conceivability in play here, one is Kripke's sense in which the sentence

(6) Water is not H_2O.

is conceivable but impossible, and another sense in which the sentence is not even conceivable. This requires the distinction above. The sentence (6) is primarily conceivable but not secondarily conceivable. This distinction corresponds to two ways of considering possibilities which we have previously contemplated: primary possibility is tantamount to considering a world as actual, while secondary possibility is tantamount to consider a world as counterfactual. Primary conceivability is tied to a priori knowledge in the sense that, for all we know a priori, it is conceivable that water is not H_2O. Secondary conceivability takes into consideration empirical facts about the actual world and empirical facts about our world that exclude the possibility of water not being H_2O. This marks a different approach in understanding Kripke's analysis of the necessary a posteriori: 'water is not H_2O' cannot be ruled out a priori, so it is primarily conceivable. However, the sentence can be empirically ruled out, hence it is not secondarily conceivable since 'water' designates H_2O in every possible world.

We have now two additional senses of conceivability: on the one hand, we are evaluating what *plays the watery role*; on the other hand, we ask about the actual reference of 'water'. Combining primary and positive conceivability, 'Water is not H_2O' is *primarily positive conceivable* in the sense that we can coherently imagine a situation in which watery stuff — the liquid that fills rivers and lakes, that we drink when we are thirsty, that falls from the sky etc. — picks out something other than H_2O. But the sentence is not

secondarily conceivable if we are to inquire about the reference of 'water' in whatever counterfactual scenario, since it is always H_2O.

This is the distinction that Chalmers suggests as the way to verify which notion of conceivability is the best guide to metaphysical possibility and to account for the alleged counterexamples to the link between conceivability and possibility. Instances of the necessary a posteriori such as 'water is not H_2O' have two different intensions associated with it: they have a *contingent primary intension* and a *necessary secondary intension*. Two ways of considering the modal status of the statement correspond to two ways of considering conceivability and possibility, as previously pointed out.

(7) Water is H_2O

is primary conceivable if one can conceive of a possible world where the primary intension is true (strongly tied to a priority). (7) is secondarily conceivable if one can conceive of a world where the secondary intension is true (this is not tied to a priority, on the contrary, this is the way to represent empirical facts of the world). 'Water is not H_2O' is primary conceivable as we can conceive a XYZ-world lacking H_2O, nevertheless containing watery stuff, and it is not secondarily conceivable since the secondary intension is contingent: Taking the semantic facts of the actual world (where water is H_2O) as fixed, the statement is false at the XYZ-world.

Now, we can say that a statement is primarily possible at a world that verifies the statement (where the primary intension is true) and secondarily possible at a world where the secondary intension is true. What kind of conceivability entails possibility? Primary conceivability entails primary possibility and secondary conceivability entails secondary possibility. The link between conceivability and possibility is not controversial if we accept Chalmers's dimensions of conceivability and possibility. It is clear that:

(CP+) Ideal primary positive conceivability entails primary possibility.

(CP–) Ideal primary negative conceivability entails primary possibility.

This distinction leads to a refinement of the conceivability argument (Chalmers 2010: 148):

(P1) P&¬Q is 1-conceivable.

(P2) If P&¬Q is 1-conceivable, P&¬Q is 1-possible.

(P3) If P&¬Q is 1-possible, physicalism is false.

(C) Physicalism is false.

However, primary possibility presents no threat to physicalism, so (P3) is false. The falsity of physicalism requires the secondary possibility of P&¬Q. Primary conceivability is a safe guide to primary possibility but it is not a safe guide to secondary possibility in the examples considered so far. There is still a gap between primary possibility and secondary possibility. Chalmers proposes a unique way to close the gap between primary possibility and secondary possibility. For, there are certain cases in which primary conceivability does entail secondary possibility.

In standard cases of theoretical identities, taking (6)'water is not H_2O' to be conceivable but not possible is actually to take (6) primary conceivable, but not secondarily possible. Primary conceivability is not in general a good guide to secondary possibility. Rather, secondary conceivability is a good guide to secondary possibility and since 'water is not H_2O' is not secondarily conceivable, it cannot be secondarily possible by that inference. This explains that conceivability is usually not a guide to possibility: it is the wrong king of conceivability we are considering. However, there are some special cases in which primary conceivability entails secondary possibility.

If some linguistic expression Q has coinciding primary and secondary intensions, then the same possibilities will verify Q and satisfy Q since intensions are defined as functions from possibilities to truth-value. If Q has the same truth-value regardless of the possibility in which Q is evaluated, then there is no gap between primary and secondary possibility. In order for the sentence P&¬Q to be secondary possible, both P and Q must have coinciding primary and secondary intensions. then the same possibilities will verify Q and P.

In this sense, primary conceivability of P&¬Q will also be secondary conceivability, so primary conceivability will, after all, imply secondary possibility. Chalmers thinks that there is only a gap between primary possibility

and secondary possibility if the primary and secondary intension of the expressions in the sentence differ. If they, instead, return the same truth-value, then primary possibility and secondary possibility will also coincide.

But what kind of special cases are these in which primary conceivability entails secondary possibility? They are those involving linguistic expressions that designate phenomenal properties, hence, cases involving phenomenal terms and also microphysical terms like H_2O. It is somewhat uncontroversial that phenomenal terms have coinciding primary and secondary intensions. The reason for the stability of phenomenal terms is quite straightforward: we have evidence that primary and secondary intensions differ when the intension of the linguistic expression and its extension seem to come apart. However, when the linguistic expression designates a phenomenal property, there can be no conceiving of phenomenal properties as different from their appearance, since it is the appearance of phenomenal properties that makes them what they are. There is a clear difference between the appearance of contingency of phenomenal terms and the appearance of contingency of physical terms. For there is a potential dissociation between appearance and reality in the case of 'water' and 'heat', on the one hand, which does not occur in the case of conscious phenomena such as 'pain', on the other hand. It is clear that something can look like water, or satisfy the watery role, but not be water, but it is clearly false that something might feel like pain and fail to be pain, since pain is essentially what feels like. The two-dimensional representation of phenomenal concepts must distinguish them from other rigid designators like 'water' that have a constant secondary intension, but a variable primary intension: the secondary intension of phenomenal concepts coincide with their primary intension. Of course, the coincidence of primary and secondary intensions of any phenomenal concept is no independent evidence for the claim that it affords us insight into the very essence of experiences.

In order to close the gap between primary and secondary possibility, microphysical truths P must also be semantically stable—have coinciding primary and secondary intensions. This is more hard to argue. There is an intuitive sense in each term which designate microphysical truths like 'mass', 'H_2O' are not twin-earthable for the same reason as phenomenal

terms are not twin-earthable, these terms already designate their referents essentially. There is no way the world could turn out to be that 'H_2O' would not be H_2O. Of course, this is the opposite from rigid designators, which have a variable primary intension. However, it could be argued that P does not have coinciding intensions. The primary intension of a physical term P could be whatever plays the P role, whereas the secondary intension of P is tied to the property that *actually* plays the P-role. However, let us consider that P has coinciding primary and secondary intensions for the sake of the argument.

Now we can correctly display the formulation of the conceivability argument:

(P1) P&¬Q is 1-conceivable.

(P2) If P&¬Q is 1-conceivable, then P&¬Q is 1-possible.

(P3) P and Q have coinciding 1 and 2-intensions. (P and Q are semantically stable)

(P4) If P&¬Q is 1-possible, then it is 2-possible. (from (P1), (P2) and (P3))

(P5) If P&¬Q is 2-possible, then physicalism is false.

(C) Physicalism is false.

The fact that P and Q have coinciding primary and secondary intensions assures the entailment form primary possibility to secondary possibility. Hence physicalism is really threaten by the conceivability argument. Of course, at this point one would argue against the coincidence of primary and secondary intensions of phenomenal terms or of microphysical terms. But we will grant that for the sake of the argument.

2.5 Kripke's modal argument

Kripke's conceivability argument against physicalism follows Descartes original pattern, but it is grounded on his own views about identity and modality. Consider the identity:

(8) P=Q

where P stands for 'pain' and Q stands for a brain state like 'stimulation of c-fibers':

(P1) If P and Q are a rigid designators, then it is necessary that P is Q.

(P2) P and Q are rigid designators.

(P3) P and not Q is conceivable.

(P4) if P and not Q is conceivable, then P and not Q is possible.

(C5) P and not Q is possible. (from P3 and P4)

(C6) P is not Q. (from P1, P2 and C5) contingency

(P1) and (P2) are assured by Kripke's semantics. Q is a rigid designator by stipulation. P is a rigid designator because 'pain' cannot pick out something other than the feeling of pain; there is no way that pain can come apart from the appearance of pain, so it always will designate pain across possible worlds. According to Kripke's semantics, identity between rigid designators is always necessary. (P3) is to a certain extent uncontroversial. We can indeed conceive of pain and stimulation of c-fibers coming apart as we can conceive of the falsity of any necessary proposition provided it is a posteriori. (P4) is the key premise of conceivability arguments in general—the link between conceivability and possibility. Kripke argues against the link regarding other kinds of identity with regards, in particular, to theoretical identity statements. He argues by explaining away the appearance of contingency of those identities. However, he thinks that the appearance of contingency of psychophysical identity cannot be explained away. Because there is no asymmetry between appearance of pain and pain as there is in theoretical statements.

According to Kripke, the trick to explain away the appearance of contingency is to point out that the identity only seems contingent because of how the referents of the terms are fixed. The reference of 'water' and the reference of 'H_2O' coincide only contingently: It is a contingent fact that what has the appearance of water is H_2O.

To illustrate this, Kripke asks us to consider a qualitatively identical situation in which 'heat' does not designate 'molecular motion' and yet someone, who is in this qualitative epistemic situation would not be able to distinguish both situations. Heat would feel like heat in both situations, but since in the twin-heat scenario, there is no molecular motion, we say that there is no heat. Because we cannot qualitatively distinguish both situations, we can conceive of the identity being false.

In the psychophysical identification, in contrast, there is no gap between the qualitative epistemic situation and the actual situation. They both coincide, so the appearance of contingency remains without explanation. This is so because, contrary to the theoretical identity statements, it is not a contingent fact that pain feels like pain. Consider now an epistemic qualitatively identical scenario where there is pain but no stimulation of c-fibers. In this twin-pain scenario, there must be something else that produces the sensation of pain. Nevertheless, this is still a scenario in which there is pain. This contrasts with the twin-heat scenario, where there is the sensation of heat without molecular motion. The disanalogy in the twin-pain scenario is revealed because what plays the pain-role must be pain regardless of what pain is (stimulation of c-fibers or something else). The appearance of the phenomenal state and the phenomenal state do not come apart.

To put this in Kripke's technical terms, the reason why the strategy applied to standard theoretical identification does not work in psychophysical identification is that, in the former case, the reference of 'water' is fixed via the referent's contingent properties: the reference of 'water' is picked up by superficial and contingent features such as being watery stuff. In contrast, the reference of a phenomenal term such as 'pain' is not fixed via its contingent properties, rather, it is fixed directly by its essential immediate phenomenal property. In theoretical cases, the terms are twin-earthable: it is possible that their appearance and their nature come apart, in phenomenal case they are not. Hence we cannot explain the appearance of contingency. The reasoning is: If they *seem* to come apart and we cannot explain away their separation, then they *are* different.

Kripke's argument was originally formulated to refute the Type Identity Theory. Nevertheless, the argument can also be adapted to attack

our formulation of physicalism. We just have to substitute the first two premises of the argument and thus obtain the zombie argument. Instead of the identification thesis, we can take the psychophysical conditional P→Q. The psychophysical conditional must be necessary for physicalism to be true. If the conditional P→Q is necessary, then any essential property of P must entail an essential property of Q. Granting the possibility of the mind existing without the body requires either abandoning the necessary connection between them, or showing that the possibility of distinction is merely an appearance. According to Kripke, such explanation is not available, hence the psychophysical conditional is false.

Chalmers' and Kripke's conceivability arguments both depart from the conceivability of the distinctness of the physical and the phenomenal to arrive at the possibility of their distinction. Both arguments make the same point, in fact, the only difference between them is the justification to infer possibility from conceivability. Kripke justifies this in terms of his direct reference theory and how words have their meaning fixed, whilst Chalmers does it by means of his interpretation of two-dimensionalism.

If the two-dimensional analysis of a posteriori necessities is correct, it should also work for psychophysical identities or conditionals. If 'pain' is identical to 'stimulation of c-fibers', then it is not secondarily possible that there be a physically identical world with no pain (zombie-world). But then what are we conceiving when we conceive of zombie-worlds? We are primarily conceiving that there are zombies i.e. we are conceiving of a world where the primary intension of 'pain is stimulation of c-fibers' is false. To find this proposition, we must locate the primary intension of 'pain', and it seems that here the primary intension is something like 'the unpleasant feeling that I get when I pierce my ears' and—assuming the a posteriori identity holds–the secondary intension is the basic physical description of 'stimulation of c-fibers.' The primary intension corresponds to a priori facts about pain, whilst the secondary intension corresponds to empirical facts about pain. So, according to two-dimensional semantics, we are allowed to infer that there is a scenario in which 'that unpleasant feeling' does not pick out anything despite the fact that there is stimulation of c-fibers. This is the zombie world, since it involves the supposed physical part of pain without

the phenomenal part. Thus, since two-dimensionalism vindicates the inference from primary conceivability to secondary possibility, the conceivability of zombie worlds assures their primary possibility (at least in the phenomenal case), so physicalism as stated in the psychophysical conditional, is false.

* * *

The two physicalist reactions to the conceivability argument previously considered are: the rejection of the conceivability of zombies and the rejection of the link between conceivability and possibility. The first reaction is implausible: we have accepted that P&¬Q is conceivable, since it is not a priori deducible that ¬Q. Conceivability is a priori consistency. If P and Q are conceptually independent, it is reasonable to accept that P&¬Q is conceivable. The second option for the physicalist is to reject the link between conceivability and possibility. This route is not so trivial, for early attempts to reject this link were based on the existence of counterexamples examined previously such as the Goldbach conjecture and instances of necessary a posteriori. However, two-dimensional semantics explains why the so-called counterexamples are not counterexamples at all, since the link between conceivability and possibility is far more subtle than initially presupposed. I have showed that the conceivability argument depends on the two-dimensional semantics to succeed—especially on the distinction between different dimensions of modality and conceivability. If this dependence is correct, then we are left with the work of rejecting the generalized version of two-dimensionalism in order to break the connection between conceivability and possibility. This should be one way to go.

Another option—to be explored in the next chapter— consists of breaking the connection by mobilizing phenomenal concepts. That would require that we change our definition of physicalism. Instead of a priori physicalism, we shall explore the prospects of developing a version of a posteriori physicalism, in which the connection between phenomenal truths and physical truths is still necessary but a posteriori. The strategy that mobilizes phenomenal concepts to defend physicalism from the epistemic arguments

is called the *phenomenal concept strategy*[3]. It will be exposed and examined in Chapter Three.

3 The Knowledge argument

Another strong and widely debated argument against physicalism is the knowledge argument formulated by Frank Jackson (1982, 1986). Mary, the brilliant scientist, has learned everything there is to know about colors and physiology of colors. She has been kept captive in a black and white room without ever having the possibility of seeing colors. As she is finally released from the black-and-white prison, she sees for the first time a red object, say, a red rose. When confronted for the first time with a red rose, she seems to learn something new, something about what it is like to see red. Mary seems to learn a new *fact* about color experience: she now knows *what it is like to see red*. But she already knew the complete physical truth about color experiences during the time of her imprisonment. If she learns a new fact, then this new fact must be non-physical. Taking this into account, this would render physicalism as false. The knowledge argument is originally formulated in terms of physical information rather than knowledge of facts. The knowledge argument, as first presented by Jackson (1982), could be formulated so:

(P1) Mary has all the physical information concerning human color vision before her release.

(P2) But there is some information about human color vision that she does not have before her release.

(C) Therefore, not all information is physical information.

Horgan (1984) considers that to formulate the argument in terms of information is somewhat ambiguous since one could make an epistemological

[3]The label is coined by Daniel Stoljar (2005).

and ontological reading. To disambiguate, Nida-Rümelin (2002) makes two possible readings of the argument explicit by formulating two versions of the argument. She proposes that 'physical information' be replaced by (i) 'complete physical knowledge about x' on the *epistemological reading* and (ii) by 'to know all the physical facts about x' on the *ontological reading*. These formulations lead to different conclusions. Consider the two versions of the argument (Nida-Rümelin 2002):

Epistemic version:

1a. Mary has complete physical knowledge concerning facts about human color vision before her release.

2a. But there is some kind of knowledge concerning facts about human color vision that she does not have before her release.

3a. Therefore, there is some kind of knowledge concerning facts about human color vision that is non-physical knowledge.

Ontological Version:

1b. Mary knows all the physical facts concerning human color vision before her release.

2b. But there are some facts about human color vision that Mary does not know before her release.

3b. Hence, there are non-physical facts concerning human color vision.

In order for the knowledge argument to be a challenge to physicalism, one needs to formulate the stronger version of the argument, the ontological version. Only the ontological conclusion is incompatible with physicalism. In its epistemic version, the argument does not succeed in providing an anti-physicalist conclusion. In (3a), it is claimed that there is some incomplete knowledge concerning some physical fact. Conversely, 'incomplete knowledge' does not mean that there is a fact missing. Examples of the sort of incomplete knowledge of facts involve indexical knowledge. Let us say I know that Julia is in Frankfurt, but a self-locating belief is missing here: I

know that Julia is in Frankfurt, but I do not know that I am in Frankfurt, for I have forgotten whom I am. When I learn that I am in Frankfurt I will not have learned a new fact, but I will have learned the same old fact under a different guise. This is how some physicalists will react to the knowledge argument. They will accept the weaker version of the argument and mobilize cases of indexical beliefs. This can be observed in the previous example by analogy applying it with phenomenal cases. The real threat to physicalism is made by the ontological conclusion that claims the existence of new non-physical facts. The kind of responses that be will explored in the following chapters will focus on this step of the argument. To make the argument more explicit in all of its readings, Nida-Rümelin formulates the ontological version of knowledge argument that embedded the epistemic formulation:

(P1) Mary has complete physical knowledge about human color vision before her release.

(C1) Therefore, Mary knows all the physical facts about human color vision before her release.

(P2) There is some knowledge concerning facts about human color vision that Mary does not have before her release.

(C2) Therefore (from (P2)), there are some facts about human color vision that Mary does not know before her release.

(C3) Therefore (from (C1) and (C2)), there are non-physical facts about human color vision.

How exactly does the knowledge argument violate the thesis of minimal physicalism? Roughly, the idea is that one might know all the objective, physical facts about the phenomenal realm, like color experiences, and yet fail to know certain facts from a subjective point of view, for instance, what is like to be in those phenomenal states; therefore, there are facts about our experiences that fall out of the physicalist's scope, so physicalism is false.

The thesis of minimal physicalism states that any truth in our world is entailed by the physical truths of our world. This includes truths about

phenomenal experiences. Let P be a complex sentence in some idealized language that describes the complete physical story about our world. An essential step in Mary's argument is the link between deducibility and necessitation: One can say that P entails Q if and only if Q is a priori deducible from P. This is precisely the opposite of Mary's case. Mary knows all the physical facts about colors (knows P) but cannot deduce all phenomenal truths (Q) from her complete knowledge of physical facts. Hence, there is no a priori connection between P and Q, as it is required by the physicalist. Again, like in the conceivability argument, the argument's most important step is the premise that if there is no a priori connection between physical facts and phenomenal facts, then there is no necessary connection. Physicalists must claim that the connection is necessary or physicalism is false. The knowledge argument is valid. So the physicalist needs to argue for the falsity of one of the premises to block the anti-physicalist conclusion (C3). There are two kinds of physicalist responses to the knowledge argument: (i) There are those who reject Mary's apparent epistemic progress and (ii) those who grant Mary's epistemic progress, but reject the ontological inferences proposed by the argument.

3.1 No epistemic novelty

The first physicalist reaction to the argument is to deny (P2), that is, to deny that there is any novelty (any kind of knowledge—factual or not) when Mary leaves the black and white room. Daniel Dennett (1991), for example, denies the qualitative character of consciousness and, consequently, the idea that qualia might be objects of knowledge. According to Dennett, there is no epistemic gap between her knowledge prior and subsequent to her release. In fact, Dennett reasons that Mary, upon seeing colors for the first time, would not be surprised at all. She would definitely recognize whatever kind of experience she did not have before as an experience of a certain kind, which she knew before under physical descriptions. Moreover, Dennett claims that our feeling that Mary learns something new is due to the fact that it is very difficult to imagine what it is like for someone to acquire all physical knowledge even from a strict field of physics (such as the physics of

visual experiences). This difficulty may not rise from the sheer quantity of so much knowledge, but from the possibility that the notion of acquiring *all* knowledge does not even make sense for our essentially limited, incomplete scientific world-view. He fabricates a new ending for Mary's story in order to make his point clear:

> And so, one day, Mary's captors decided it was time for her to see colors. As a trick, they prepared a bright blue banana to present as her first color experience ever. Mary took one look at it and said 'Hey! You tried to trick me! Bananas are yellow, but this one is blue!' Her captors were dumfounded. How did she do it? Simple,' she replied. 'You have to remember that I know everything-absolutely everything-that could ever be known about the physical causes and effects of color vision. So of course before you brought the banana in, I had already written down, in exquisite detail, exactly what physical impression a yellow object or a blue object (or a green object,etc.) would make on my nervous system. So I already knew exactly what thoughts I would have. (...) I realize it is hard for you to imagine that I could know so much about my reactive dispositions that the way blue affected me came as no surprise. Of course it's hard for you to imagine. It's hard for anyone to imagine the consequences of someone knowing absolutely everything physical about anything.' (Dennett 1991: 399)

We can observe that Dennett appeals to the truth of (P1); Mary simply knows everything there is to know about color physiology, therefore there is nothing physical left for her to learn. But holding physicalism true makes the first premise inconsistent with a premise that posits any epistemic progress given complete knowledge of the physical facts. Dennett does not provide any definite arguments, nonetheless he confronts one intuition pump (that Mary learns something when she leaves the room) with another intuition pump—that she does not. In fact, Dennett is not alone in his thinking. Remarkably, Frank Jackson (2004), on reconsidering the case, is an exponent of this kind of response to the knowledge argument:

> In some worlds, its [qualia] nature cannot be deduced in principle from the full account of the physical nature of that world, but in

> other worlds, including ours, it can. The redness of *our* red can be deduced in principle from enough about the physical nature of our world despite the manifest appearance to the contrary that the knowledge argument trades on. This is why I now think that the knowledge argument fails. (Jackson 2004: 417)

This kind of response is popular enough but leads to an old problem given that is confronts two contrary intuition pumps. Let us therefore proceed with the intuition that Mary does learn something new and explore other responses the physicalist may have.

3.2 Some epistemic novelty

If we grant that Mary gains knowledge upon leaving the black and white room, the question arises as to what kind of knowledge she gains. Proponents of the knowledge argument require that Mary's epistemic gain corresponds to knowledge of new facts—this is why physicalism would be refuted by the knowledge argument: Mary gains knowledge of *non-physical facts*, i.e. true propositions outside the realm of physics. In order to block the anti-physicalist conclusion one needs to argue that Mary gains non-propositional knowledge. What kind of non-propositional knowledge? We shall now examine two possibilities: the ability hypothesis and the new mode of presentation (Fregean) strategy.

3.2.1 Ability Hypothesis

Another way to contest the knowledge argument is by rejecting (P2). According to the ability hypothesis of Nemirow and Lewis, after Mary's release, she does makes *some* epistemic progress ((P2) is true). However, she gains no propositional knowledge. This is substantiated in our feeling that Mary does not gain knowledge of facts ((C2) is false). The task of the ability hypothesis is to provide an error theory for our feeling that Mary gains propositional knowledge upon her release. That feeling leads us to accept the epistemic step in the argument present in (C2) and interpret it as propositional knowledge.

The ability hypothesis claims that Mary's epistemic gain should be interpreted as the acquisition of an ability rather than the acquisition of a

new fact. If there is no fact to be learned, the argument fails. According to Nemirow (1980, 1990), this response to the knowledge argument consists of distinguishing between different modes of understanding: not grasping of facts, but in the acquisition of abilities. This shall reflect different notions of knowledge, which were originally distinguished by Gilbert Ryle (1949). The general idea is that the epistemic gain, obtained from gaining abilities, is non-propositional. In contrast, knowledge of facts is propositional knowledge, and therefore, eliminates possibilities.

The abilities to which they refer are abilities to 'place oneself in a state representative of the experience' (Lewis 1994). Exercising that ability is adopting the point of view of the subject of the experience. This is a way to explain the novelty in Mary's experience without positing extra subjective facts. One gains new abilities upon undergoing some relevant experiences that enables the experiencer to recall this experience, discriminate among similar experiences, recognize instances of the same experiences. The ability hypothesis proposes to interpret knowing what it is like, as it occurs in the knowledge argument, as the acquisition of certain abilities.

3.2.2 Fregean strategy

A third option for the physicalist is also within the group of responses that accepts (P2), but rejects (C2). Like the ability hypothesis, they grant that Mary gains some kind of knowledge, but they deny that she gains knowledge of new facts. The difference between this set of responses to the ability hypothesis response is in specifying the kind of knowledge Mary gains. Previously, Lewis and Nemirow called it know-how or acquisition of abilities. Another group of philosophers consider the possibility that Mary gains knowledge of a new mode of presentation of an old fact she already knew while confined in the black and white room, except that this old fact was known to Mary through its physical mode of presentation. She acquires now a broader conception of a fact she already knew while inside the black and white room, or, as proponents of this strategy say it: this is the old fact, new guise strategy. That is roughly the same Fregean strategy used to explain cognitive significance in coreferential terms like Hesperus and Phosphorus.

The strategy under discussion now requires that the content of Mary's knowledge be finely individuated. The so-called coarse-grained account of propositions is represented by the standard semantics in terms of possible worlds. In this framework, propositions are just sets of possible worlds. Then, if Mary knows that P, then she has eliminated from her epistemic horizon all possibilities 'outside' of, i.e. incompatible with the proposition P. The problem with this coarse-grained account of propositions is that it does not work very well in epistemic contexts. Consider examples from algebra, the sentence '1+1=2' is true in all possible worlds (for it is necessary), but so is the sentence '5+7=12' or even the necessary a posteriori truth 'water is H_2O' and according to Kripke's theory of rigid designation, all identities, since they are all necessary. We cannot explain, within a coarse-grained account of proposition, the difference that emerges in those situations, including prominently epistemic contexts, where I may believe that X but not that X*, given that I have X and X* under different conceptualizations, or different modes of presentation. Coarse-grained individuation of mental content leads to undesirable results: identifying propositions with sets of possible worlds turns out to endorse the conclusion that there is only one necessary proposition, viz. the set of all possible worlds. In epistemic context, this way of individuating propositions leads to thesis that any subject who has contradictory beliefs is irrational, even if the subject is justified in thinking so (like Lois Lane's belief that Superman can fly and that Clark Kent cannot fly). A finely-grained account of content is needed to explain those epistemic and modal phenomena. This will be the topic of the next chapter.

The phenomenal concept strategy mobilizes the finely-grained individuation of contents. This kind of response will claim that the problem with the knowledge argument is that it seems to presuppose a coarse-grained account of propositions from the beginning, and this will lead, inevitably, to the anti-physicalist conclusion: the positing of new subjective facts. Once we distinguish between the coarse-grained account and fine-grained accounts of propositions we see that Mary's new knowledge may not exclude any possibilities, thus not constituting knowledge of new proposition in the coarse sense.

4 Systematizing the reactions to the anti-physicalist arguments

We can now systemize the anti-physicalist arguments pointing out their common structure. Both arguments, the conceivability and the knowledge argument, consist of two steps that allegedly lead to the falsity of physicalism, as explained above. The first is the epistemic step; in Mary's case, the gap is formulated in terms of a priori deducibility: phenomenal knowledge cannot be deduced a priori from physical knowledge. In the zombie case, the gap is formulated in terms of conceivability: creatures physically identical to us but phenomenally different are conceivable. The second step is an inference from the epistemic gap to a metaphysical gap. In Mary's case, from the non-deducibility claim, it is inferred that there are non-physical facts: In the zombie case, the inference goes from conceiving of x, to x is metaphysically possible. The third step is the contradiction with physicalism: the existence of non-physical facts and the metaphysical possibility of zombies contradict physicalism. Both arguments follow a similar structure. In fact, formulating in terms of deducibility is equivalent to formulating it in terms of conceivability. In Jackson's argument, Mary cannot deduce phenomenal truths from physical truths, hence phenomenal truths are not metaphysically necessitated by physical truths, making physicalism false. So it can be concluded that non-deducibility implies non-necessitation. In Chalmers' argument, conceiving P without Q is equivalent to claim that Q is not deducible from P. If Q is not deducible from P, then P does not metaphysically necessitate Q, hence physicalism is false. Both arguments then start from epistemic premises to proceed to a metaphysical conclusion.

We have seen that there are at least two paradigmatic ways of responding to these arguments. Some physicalists, coined as type-A materialists, reject the epistemic step. They claim that the epistemic gap is merely an illusion that derives from misunderstandings regarding the definition of the physicalist thesis. In their view, Mary does not gain genuine knowledge at all as she is released from the black-and-white room, and zombies are simply inconceivable. The second line of response, advanced by type-B materialists, emanating from moderate physicalists, recognizes the existence of

an epistemic gap but rejects the inference to the metaphysical gap. Thus, physicalists deny the link between conceivability and possibility. It means that Mary gains phenomenal truths that she cannot a priori deduce from her previous knowledge. But this does not present a threat to the physicalist anymore, for the physicalist step back from any commitments to a priori reductive explanation. Moreover, zombies are conceivable, but metaphysically impossible. The more moderate branch of physicalists hold that the epistemic gap is a consequence of the way we think about consciousness. They recognize the gap but do not think it corresponds to an ontological gap. There is no gap between physical processes and conscious processes. The gap is at the level of concepts. The psychophysical conditional

(PHYS) $P \rightarrow Q$

must be posteriori in order to accommodate the first step of the conceivability and the knowledge argument and the truth of physicalism. The epistemic gap, which can be understood as an inferential disconnection between P and Q should be explained as a consequence of the difference between phenomenal concepts and physical concepts.

Chapter 3

The Phenomenal concept strategy

1 Terminology: Concepts and contents

Content is an abstract entity associated with concrete entities, such as utterances or sentences yielding the linguistic content of an utterance or a thought having mental content. Content of utterances or thoughts is usually expressed by that-clauses embedded in indicative sentences. The content of the thought expressed by the embedded sentence in:

(1) Claudius thinks that it is sunny in Lisbon.

is what appears on the right hand side of the relative pronoun 'that': it is sunny in Lisbon. Thoughts, on the other hand, are constituted by concepts. Concepts are mental representations that compose thoughts.

In the *phenomenal concept strategy*, concepts and contents are fine-grainedly individuated. The fine-grained individuation, as we have seen in the previous chapter, allows that two different concepts expressing different contents refer to the same entities. Thus, ***Hesperus*** and ***Phosphorus*** are different concepts that refer to same entity: the planet Venus. The content of a thought involving the concept ***Phosphorus*** may be different from the content of a thought involving the concept ***Hesperus*** even though they denote the same entity. The two thoughts may play different roles e.g. in belief

ascriptions. As a result, if a concept is fine-grainedly individuated, a person can hold contradictory beliefs about the coreferential extensions without being charged with irrationality. A fine-grainedly individuated concept can be substituted by other co-referring concepts in *intensional context* preserving the truth-value of the sentence, whereas a coarse-grainedly individuated concept cannot be substituted by *extensional context*.

Understanding ideas, words, recognizing objects, judging speeches, evaluating our own feelings can only be done if there is some degree of conceptualization. There are certain concepts that occur in the ascription of Claudius' thought 'It is sunny in Lisbon' e.g., ***Sunny***, ***Lisbon***. Claudius needs to employ these concepts in order to form this thought. If Claudius did not possess some of these concepts he would not be able to recognize that Lisbon is sunny and so on, he would be *conceptually blind* to this fact. Conceptualization is required for us to judge, recognize, discriminate things and thoughts. For example, a child that possesses no concept of numbers might interact with *three* different toys without recognizing that there are three different toys in front of it. G.H. Wells, for instance, tells a story about a lost mountaineer in the Andes who wakes up in a country where the inhabitants are all blind. For generations, the dwellers of this country are not able to see as a result of a disease that afflicted the village hundred of years before. They have no concept of visual abilities. They simply cannot understand or imagine what the traveller wants to tell them when he mentions colors, when he says he is *seeing* the path instead of *hearing* the path, when he talks about light and darkness. There is no way of conceiving a certain state of affairs without the employment of the appropriate concepts.

2 Phenomenal concepts

Qualia constitute the objects of introspective knowledge: the subject matter of which we can form, justify and eventually verify beliefs. Knowledge requires belief, and belief requires concepts. Thus, the natural assumption that follows from the thesis of the existence of qualia is that there are special phenomenal concepts, whose application would allow the discrimination of mental states and their qualitative aspects. Brian Loar introduces his

seminal theory of phenomenal states as follows:

> On a natural view of ourselves, we introspectively discriminate our own experiences and thereby form conceptions of their qualities, both salient and subtle. [...] What we apparently discern are ways experiences differ and resemble each other with respect to what it is like to have them. Following common usage, I will call these experiential resemblances phenomenal qualities, and the conceptions we have of them, phenomenal concepts. Phenomenal concepts are formed 'from one's own case'. (Loar 1997: 597)

Phenomenal concepts derive from our conscious experience which we access introspectively. Some degree of conceptualization regarding one's own experience is required to attend to it, otherwise one would be blind to one's own feelings just like small children are conceptually blind to numbers. The phenomenal concept ***red sensation***$_P$ is the concept of the specific type of sensation someone typically has when looking at red things. It is very different from the concept ***red***, for this concept refers to red things. It is also different from the concept ***the sensation caused by red things*** for certain color-blinds will feel green upon looking at red things (Stoljar 2005).

It is also important to differentiate phenomenal concepts from psychological concepts (Balog 2009). The latter are analyzed in functional terms, that is, in terms of the causal-functional relation that the psychological state bears with other states. Consider the psychological concept ***sweetness*** and the phenomenal concept ***sweetness***$_P$. The reference of the psychological concept ***sweetness*** can be characterized as the state that is caused by the contact between sugared foods and the taste buds in normal conditions. Whereas the referent of ***sweetness***$_P$ picks out a certain sensation (sweetness) directly, without a functional mode of presentation. Physicalists assume that, psychological and phenomenal concepts refer to the same physical properties, although in different ways. A psychological concept is an objective, third-person concept, since it refers to a mental state, which contains no reference to the subject's subjective experience. In fact, a blind person may acquire a certain psychological concept of color from a third-person point of view, that is, by testimony, whilst a phenomenal concept is

formed exclusively from 'one's own case'.

Some physicalists think that appealing to phenomenal concepts is the most promising strategy to disarm anti-physicalist arguments. The phenomenal concept strategy aims at providing an alternative physicalist explanation of the epistemic gap posed by anti-physicalist arguments, that is, of the conceptual independence of physical and phenomenal concepts. This alternative explanation is designed so as to not expand the ontology beyond physical facts. The *sui generis* status of phenomenal concepts is due to its experience dependence and not due to their referring to special realm of non-physical properties. Phenomenal concept theorists ascribe to concepts of our experience unique aspects that distinguish phenomenal concepts from other concepts, but they still refer to physical properties. One candidate criterion to separate phenomenal concepts from physical concepts is offered by the so-called *experience thesis*, the thesis that *possession* of these concepts are experience-dependent:

> *Experience Thesis*:
> One can only *possess* a phenomenal concept C, if one undergoes the relevant corresponding experience.

The experience thesis is a thesis about concept possession. However, many versions of the strategy do not endorse the experience thesis in the sense of being a requirement for concept possession. Some phenomenal concept theorists think that experience is a requirement for special reference-fixing mechanisms, for concept acquisition or even for involving different psychological faculties. Regardless of which condition each version of the strategy thinks it is satisfied by experience, *all* versions agree that phenomenal concepts are experience-dependent, and that they are, per definition, perspectival. I will begin by exploring the experience thesis, for it seems to be a general commitment shared by many phenomenal concept theorists. I shall return to accounts that focus on the *special reference-fixing mechanism* of phenomenal concepts. Papineau (2007) writes:

> The crucial feature of phenomenal concepts [...] is that they are experience-dependent: the concept's acquisition depends on its

> possessor having previously undergone the experience it refers to. (Papineau 2007: 126-127)

Papineau's passage highlights the importance of special acquisition's conditions of phenomenal concepts. Indeed, this aspect figures as a widespread commitment of proponents of the strategy. Unpacking the experience thesis as a guide to the conceptual independence of phenomenal concepts will help us to make the case for the phenomenal concept strategy. At this point, we are interested in the constraints on phenomenal concepts in order to yield a satisfactory physicalist response to the anti-physicalist arguments of the sort explored in the previous chapter. Then, we shall critically assess the constraints that are usually offered to characterize phenomenal concepts. At last, we shall explore versions of the strategy that are not committed to experience as a condition for concept possession. Instead, these other versions focus on experience as a condition for special reference-fixing mechanisms of phenomenal concepts.

Many proponents of the strategy seek to establish the existence of a binding relationship between certain type of experience and the possession of phenomenal concepts. They claim that the *only way* one might possess phenomenal concepts is by undergoing an appropriate experience. In Lewis' words: in some cases experience is the best teacher. Indeed, the phenomenal concept theorist wants to say that, in this case, experience is the *only* teacher. Some items of knowledge cannot be grasped unless the knower has experienced first-hand the content of concepts involved in those items of knowledge. Objective physical descriptions of these experiences are not adequate *ersatz* for the experience in the first-person perspective. There is something very intuitive about this. In a sense, it is a trivial claim that phenomenal concepts are tied to experience. Phenomenal concepts are concepts *about* experiences formed from one's own perspective: in order for someone to attend to their own experience one needs to possess the appropriate concept, in the same way as one needs to possess appropriate concepts in order to understand thoughts, ideas etc.. It is natural to strengthen this to the thesis that phenomenal concepts are constitutively tied to experiences, that is, experience is a necessary condition for one to come in possession of phe-

nomenal concepts. If phenomenal concepts are special because they have special possession conditions, then the strategy might mobilize this characteristic of phenomenal concepts to neutralize the epistemic arguments.

The phenomenal concept strategy consists in explaining the non-deducibility of phenomenal truths from physical truths due to the lack of appropriate concepts to make the appropriate deductions: like the child who can see the toys in front of it but who does not notice that there are 'three' toys in front of it because it lacks the appropriate number-concept. Or the blind man who cannot imagine what the traveller means when he talks about 'day' and 'night', because he lacks the requisite visual concepts. So the characters in our anti-physicalist arguments (Mary and the zombie) lack the relevant concepts that would allow them to deduce certain phenomenal truths from physical truths. As it happens, there is only one way to acquire such concepts: to undergo the appropriate experience. Since they never had those experiences, they cannot make the appropriate deductions.

At this point we seem to have a relevant constraint for phenomenal concepts:

The concept $\boldsymbol{C}$ is a *phenomenal concept* if and only if

1. it is conceptually independent of physical concepts;
2. it refers to phenomenal properties;
3. it is experience-dependent.

However, these constraints do not really constitute a physicalist account of phenomenal concepts, since they are jointly compatible with a dualist ontology. In fact, proponents of the knowledge argument and of the conceivability argument may subscribe to this conception of phenomenal concepts. To turn these phenomenal concept criteria into a physicalistic account of phenomenal concepts, one must strengthen the first criterion by adding that phenomenal properties are identical or necessitated by physical properties.

The general desiderata for phenomenal concepts include the following items:

1. Phenomenal concepts are conceptually isolated from physical concepts.

2. Phenomenal concepts and physical concepts are distinct *but co-referential*.

3. Phenomenal concepts are experience-dependent.

Two of these constraints are quite uncontroversial among the advocates of the phenomenal concept strategy: the co-reference of phenomenal concepts and physical concepts and their conceptual independence. However, the experience requirement is what usually generates most controversy. Tye (2009) adduces counter-examples to the experience thesis involving situations in which $\boldsymbol{Red}_P$ (i.e. phenomenal concept of red) could be produced without the presence of the experience of seeing red. Instead, what could enable us to possess such phenomenal concept $\boldsymbol{Red}_P$ would be, say, a miracle, a magic spell or even neurosurgery. These counterexamples would require for us to modify the phenomenal concept criteria we had so far. Thus, following Ball (2009: 938), we may weaken the phenomenal concept constraints to accommodate such cases as follows.

Refinement of phenomenal concept criteria:

(a) There is some phenomenal experience type e, and some property p, such that experience tokens fall under e in virtue of their relation to p.

(b) $\boldsymbol{C}$ refers to p.

(c) *Under normal circumstances*, a human being can possess C only if she has had an experience of type e.

Restricting the experience thesis to 'normal circumstances' rules out the possibility of someone forming phenomenal concepts through non-traditional (highly implausible) methods, such as neurosurgery, spells and miracles. As pointed out before, there is a sense in which (c) is absolutely trivial. It is plausible to assume that, in the obvious sense, a concept about conscious experience requires conscious experience. When it comes to taste it is very hard to conceive of a situation of knowing the taste of marzipan, for example, without ever having tasted marzipan. Of course, there are ways to miss the point, as Lewis (2004) observes: one can try to know the taste of

marzipan by testimony, that is, by having someone else explain what known taste are similar to the taste of marzipan. Marzipan is made of sugar and almond paste, I am already familiar with the taste of sugar and the taste of almond paste. Does that make me familiar with the taste of marzipan? Perhaps an experienced pastry chef (who, for some reason, has never tasted marzipan) would be able to deduce the taste of marzipan through the descriptions of the ingredients involved in the making of marzipan. Perhaps, Beethoven was able to vividly imagine the symphony he composed when he was already deaf. In both cases, as Lewis points out, the new experiences are a 'bunch of pieces of old experiences rearranged' (Lewis 2004). Rearranging old experiences does not enable us to form relevant phenomenal concepts. One thing is resembling experiences and another thing is rearranging them. The proponent of the experience thesis can always say that the result generated by rearranging or reassembling old experiences to form new experiences is relevantly different from the result of properly having a new experience: tasting marzipan and hearing the 9th Symphony. The triangulation required by this way of acquiring 'new experiences' does not result in having the experience, and even though one can imagine new experiences, it is still different from experiencing them.[1] We need to be more careful when we postulate the experience thesis as pinpointing the essential property of phenomenal concept. As mentioned earlier in this chapter, a necessary requirement for the acquisition or possession of phenomenal concepts is to undergo a relevant experience. So possessing the phenomenal concept $\boldsymbol{Red}_P$ requires having experienced red, although possessing the mere physical concept or psychological concept ***red*** does not require experiencing red; one can know by testimony that red is the typical color of fire hydrants in England, without ever having seen colors.

Even so, the last two constraints (b) and (c) are still perfectly compatible with dualism. The conceptual isolation of phenomenal concepts is usually

[1]Brian Loar (1990, 1997) is the exception here. He thinks that phenomenal concepts can be formed by means of triangulation. I will return to this point later in this chapter.

mobilized by dualists as a reason to conclude that the properties to which the concepts refer are, in fact, distinct. This is the core dualist intuition. Many proponents of the anti-physicalist arguments think that these constraints only strengthen the case against physicalism. The fact that there is no a priori connection between physical and phenomenal concepts means that they must refer to different properties. Naturally, we need to ask about the possibility of a posteriori identification. The dualist will not accept this proposal because the knowledge argument depends on an assumption to succeed: the isolation of phenomenal conceptual or, as we have formulated earlier, the non-deducibility of phenomenal truths from physical truths. This is both agreed by dualists and physicalists. The reasoning is this: phenomenal and physical concepts are conceptually independent, phenomenal concepts conceive their referents essentially, while physical concepts conceive their referents contingently. The dualist thinks that this should imply that we have a priori access to phenomenal properties. But being in pain does not grant us a priori access to the stimulation of c-fibers. So, the dualist argues, phenomenal concepts and physical concepts cannot co-refer. Because there is no a priori connection between P and Q, and Q conceives its properties essentially, the appearance of contingency cannot be explained away. So, it is not a mere appearance, the identification is contingent, hence physical concepts and phenomenal concepts do not co-refer. The conceivability argument also depends on the mentioned assumption and a further one: non-deducibility leads to non-necessitation, i.e. if Q is not deducible from P, then P$\rightarrow$Q cannot be necessary.

Physicalists that advocate for phenomenal concepts usually accept the conceptual independence assumption but reject the inference to property distinction. They have to explain the source of the conceptual isolation in a way that is compatible with physicalism. Moreover, they need to provide an explanation for the necessity of the psychophysical conditional or identity granting that P and not Q is conceivable.

The task of the phenomenal concept strategy is to answer two questions:

(Q1) Why are phenomenal concepts conceptually independent of physical concepts?

(Q2) How can the psychophysical identification or the conditional be a posteriori if phenomenal terms are semantically stable?

The answer to (Q1) usually mobilizes the *sui generis* character of phenomenal concepts, e.g. experience-dependence, acquaintance, or as we shall discuss below, a characteristic reference-fixing mechanism. It is the answer to (Q2) that will break the link between conceivability and possibility established by the two-dimensional framework as discussed in the previous chapter. The two-dimensional argument against physicalism claims that identity statements can only be a posteriori if the terms involved in the statements are *semantically unstable*. Let us recall that this is equivalent to saying that terms have different primary and secondary intensions. The assumption of the conceivability argument is that a priority and necessity only come apart in cases where we connect semantically unstable concepts. This is because with semantically unstable concepts we need empirical information to know their referent, and since empirical information is contingent, they could turn out to be different. In contrast, phenomenal terms are transparent in the sense of being stable: they we do not need empirical information to know the referent of 'pain'; what fixes their reference is, in fact, being in pain. This is also to say that they are not twin-earthable, since there is no scenario where a twin-pain is different from pain. Since phenomenal concepts are semantically stable terms, identities involving them will be necessary only if they are a priori:

> The underlying anti-physicalist thought, recall, was that semantic stability goes hand in hand with knowledge of real essences; conversely, if thinkers are ignorant of real essences, they must be using unstable concepts. The complaint about Type-B physicalism, then, is that it requires the possessors of phenomenal concepts like pain to be ignorant of the real physical essence of pain, even though the concept pain is manifestly stable. The anti-physicalists thence conclude that pain must refer to something non-physical, something with which the possessors of the concept are indeed directly acquainted. (Papineau 2007: 131)

Now let us see how some specific accounts of phenomenal concepts answer the anti-physicalist arguments discussed in the previous chapter and to the

required questions:

(Q1) Why are phenomenal concepts conceptually independent of physical concepts?

(Q2) How can the psychophysical identification or conditional be a posteriori if phenomenal terms are semantically stable?

3 Specific accounts of phenomenal concepts

The general strategy using phenomenal concepts, as I hope to have made clear, is to ascribe to them special features that make them essentially different from physical concepts blocking any possibility of deducibility between them. Although there is widespread agreement about the way of proceeding in that phenomenal concepts are experience-dependence, there is still great deal of divergence regarding the nature of phenomenal concepts, that is, what makes phenomenal concepts experience-dependent?

The phenomenal concept strategy challenges the inference from conceivability to possibility. The strategy accepts the two-dimensional treatment of standard a posteriori necessities, but rejects its application to statements involving phenomenal concepts. In other words, phenomenal concept theorists accept that certain identity statements are conceivable as false but metaphysically impossible, since they might involve semantically unstable terms. Nevertheless, they hold that this assumption cannot be applied to identity statements involving phenomenal concepts only because they are semantically stable.

There are numerous proposals for a physicalist account of phenomenal concepts. I shall expose and assess only these that are considered to be the most relevant accounts in the literature, viz. the recognitional account, the indexical account and the constitutional account of phenomenal concepts.

3.1 The recognitional account

In two seminal papers, Loar (1990, 1997) presents the first clearly articulated physicalist account of the phenomenal concept strategy. His version, as all

other subsequent versions, requires that phenomenal concepts be experience-dependent. However, the special character of phenomenal concepts is not constitutively tied to experience as is the case with other versions of the strategy. Loar's recognitional account of phenomenal concepts aims at rescuing not the psychophysical conditional from the anti-physicalist arguments but the psychophysical identity statement (as in Kripke's argument), P=Q, where P stands for 'stimulation of c-fibers' and Q stands for 'pain'.

Loar's proposal considers phenomenal concepts to be part of a wide class of concepts called *recognitional concepts* involved in the subject's recognition abilities to discriminate, classify and recognize objects:

> Suppose you go into the California desert and spot a succulent never seen before. You become adept at recognizing instances, and gain a recognitional command of their kind, without a name for it; you are disposed to identify positive and negative instances and thereby pick out a kind (Loar 1997: 600).

A recognitional concept is formed in virtue of the subject's abilities to discriminate and to re-identify the same kind of object. Recognitional concepts work like type-demonstrative concepts, they enable our ability to identify, discriminate, classify, perceive an object of *that* kind without the mediation of any description. The reason why recognitional concepts are like *type*-demonstratives rather than *token*-demonstratives is that the latter only denote particular experiences. They, thus, do not serve the purpose of identifying kinds of experiences or recalling a certain type of experience introspectively. The task of discriminating and classifying our experiences requires the ability to recall old experiences and refer to them as general experiences, instead of particular occurrences of them. When I describe to my doctor the headache that I get at night, I want to convey a recurrent type of pain and not an isolated occurrence of pain. I want to identify a recurrent headache that afflicts me at night, which requires the ability to re-identify my experience: 'that sharp pain again', or 'that same pulsing pain' etc.. Token demonstrative concepts do not convey kinds of pain, rather they convey only the 'experience I am having right now', the concept disappears when I stop experiencing it.

Recognitional concepts and physical concepts play a different role in our rational and practical system. We use recognitional concepts in our perceptual cognition to pick out its referent without any descriptive mediation. It is a sort of primitive capacity which is shared by other non-human animals. Physical concepts compose theoretical knowledge which constitutes our abilities to describe and to explain what we observe; they involve a great deal of abstraction. This points to a reason for why those two types of concepts end up being cognitive isolated from each other. Alter and Howell (2009) suggest a speculative idea that would propose an evolutionary ground for the isolation character of phenomenal concepts: the idea is that those two sorts of concepts have evolved independently. Surely, our abilities to recognize shapes and colors came before our abilities to theorize about them. This may count as evidence for their cognitive isolation. So, this is a suggestion for how to respond to (Q1).

Loar claims that the fact that phenomenal concepts are recognitional concepts points to the reason why phenomenal concepts are conceptually isolated from physical concepts: recognitional concepts pick out their reference via a non-contingent mode of presentation, just like phenomenal concepts. In Loar's view, this answers to (Q1):

> If there are recognitional concepts that pick out physical properties not via contingent modes of presentation, they do not discriminate their references by analyzing them (even implicitly) in scientific terms. Basic recognitional abilities do not depend on or get triggered by conscious scientific analysis. If phenomenal concepts reflect basic recognitions of internal physical- functional states, they should be conceptually independent of theoretical physical-functional descriptions. (Loar 1997: 228)

There is plausibility in the thesis that phenomenal concepts are recognitional concepts. First, recognitional concepts are strongly experience-dependent. Acquiring the ability to recognize instances of certain experience requires that the subject undergoes the corresponding experience.

Second, recognitional concepts, as well as phenomenal concepts, have the same special reference-fixing mechanisms. The recognitional account must tell us how phenomenal concepts have their references fixed in a way that

is so different from physical concepts and that explains the non-deducibility between them while still allowing co-reference. The special reference-fixing mechanism should lead the way to a response to (Q2): How can P=Q be a posteriori if Q is a semantically stable term? If Q picks out its referent, in Loar's terms, via a non-contingent mode of presentation, P=Q should be a priori or false. In two-dimensional semantic terms, this should be so because Q has coinciding primary and secondary intensions, whence the conceivability of (P&¬Q) entails its primary possibility. Loar intends to block the step from primary conceivability to secondary possibility by postulating that phenomenal concepts have different reference-fixing mechanisms. What is, in Loar's view, the reference-fixing mechanism of a phenomenal concept? The concept of a type $\boldsymbol{Red}_P$ is associated with a disposition to identify a token of the type $\boldsymbol{Red}_P$. The phenomenal concept of the type $\boldsymbol{Red}_P$ is associated with dispositions to discriminate tokens of red experiences from other tokens of red experiences. Loar thinks that phenomenal concepts of type experiences incorporate the tokens of that experience.

This way of explaining the reference-fixing mechanism by appealing to the incorporation of phenomenal tokens into phenomenal types is certainly insufficient. The samples of experience, i.e. phenomenal tokens incorporated in phenomenal types cannot fix any reference because a single phenomenal token exemplifies too many tokens of different phenomenal types, making it impossible to determinate which phenomenal properties are being fixed. An experience of red, for example, is a token of, not only the general type red, but also of its specific shade of red, say, crimson. For this reason, phenomenal tokens do not determine which phenomenal quality is individuated by a phenomenal concept. What determines whether a concept will refer to a specific token of experience is the fact that the concept is associated with the disposition to identify a specific token as an instance of a particular phenomenal property. The way the reference of phenomenal concepts is fixed in Loar's account is *by their association with specific recognitional dispositions*. This leads to the recognitional account thesis: *phenomenal concepts are individuated by our recognitional dispositions*.

In sum, the recognitional account proposes that a phenomenal concepts is a sort of recognitional concept since both are experience-dependent and

have special reference-fixing mechanisms, which is the fact that they are individuated by our recognitional dispositions.

Loar can now explain (Q1). Phenomenal concepts are conceptually independent of other concepts because they pick out their reference directly. They pick out their reference directly due to the associated recognitional abilities, which does not involve any conceptual mediation. Due to their recognitional character, phenomenal concepts can be acquired without any conceptual link to physical concepts. This account entails, beyond the conceptual isolation of concepts, a cognitive isolation. The cognitive isolation is meant to explain how phenomenal concepts are a posteriori connected to physical concepts (Q2):

> The conceptual property of being a posteriori must be in part psychological- cognitive, in a non-semantic sense. It is then hard to see why that psychological relation, or lack of one, should not suffice for ones concept being or not being directly inferable from the other without further premises, and hence being related a posteriori. If some such idea is adequate, it would seem to undercut the idea that we need something contingent in the semantics to explain the a posteriori status of phenomenal-physical identities. (Loar 2003: 116)

Loar's goal is to offer an alternative explanation of the semantic stability of phenomenal concepts. The assumption that proponents of the conceivability argument endorse is that phenomenal terms are stable *because* they do not depend on how the actual world is like, since their referent is fixed via non-contingent properties. So, phenomenal knowledge would be transparent in the sense that we would be able to know their referents a priori. Loar's contribution to the debate is to deliver yet an independent reason for the stability of phenomenal concept. *Because* phenomenal concepts are *not cognitively tied* to physical concepts, their identification is a posteriori. For Loar, it is not necessary to explain the conceptual independent as a result of the way they pick out their referents. It is actually sufficient to appeal to the psychological difference of phenomenal concepts and physical concepts.

Hill and McLaughlin (1999) also propose a version of the recognitional account, except that, in their view, the psychological character of phenomenal

concepts is even more explicit. They advocate that the reason phenomenal concepts and physical concepts are conceptually isolated is that the deployment of each of these concepts involves different psychological faculties due to their distinct reference-fixing mechanisms.

> It is plausible, we maintain, that the reference of the concept of pain is fixed by the fact that subjects have a commitment (or a disposition) to apply the concept to internal states that are experienced directly as having a certain qualitative feel. Further, it is plausible that the reference of (say) the concept of C-fiber stimulation is fixed by stipulation involving a description of the form 'the neural process that has such-and-such a structure and that is responsible for such-and-such experimental effects in the actual world.' Under the assumption that the reference of the two concepts in question is fixed in these very different ways, we can account for the fact that it is impossible to see a priori that the concepts have the same reference in purely psychological terms. (Hill and McLaughlin 1999: 453)

How does the recognitional account responds to the knowledge argument? In virtue of the cognitive isolation of phenomenal concepts from physical concepts, Mary is not able to a priori deduce phenomenal experience of ***Red*** from the complete knowledge of physics she had previous to her release. The step from physical descriptions to phenomenal description requires experience. Because phenomenal concepts can pick out a physical property directly and independently of physical knowledge, Mary cannot know phenomenal $\boldsymbol{Red}_P$. Thus, it does not follow that non-physical facts exist.

Regarding the conceivability arguments, explaining how the psychophysical identity can be a posteriori and necessary blocks the link between conceivability and possibility, hence it blocks the anti-physicalist conclusion of the arguments.

3.2 The indexical account

The chief exponent of the indexical account of phenomenal concepts is John Perry (2001). According to him, phenomenal concepts work like indexical concepts. We can explain Mary's new knowledge as we explain gain of

indexical knowledge in specific situations. Cases involving indexical knowledge[2] are cases of locating the content of an utterance in time and space and also the correct identification of the person that I am. Perry argues that the content of indexical knowledge is not properly captured by the dominant semantic referential theory. In the familiar Kripkean view of contents, the content of an utterance is a singular proposition containing the object/individual referred to by the indexical/proper name. This is what Perry calls 'referential content' or 'subject matter content' (Perry 2001). This kind of content raises the familiar Fregean problem, it does not allow us to distinguish cognitive states expressed by (1) and (2):

(2) I am in Frankfurt (as uttered by JT).

(3) JT is in Frankfurt (as uttered by anyone).

The utterances (2) and (3) have the same referential content: the proposition that is true if and only if JT is in Frankfurt. However, I may know that I am in Frankfurt, but as I may suffer from a serious memory loss, I may not know that I am JT and hence not know that JT is in Frankfurt. To give an account of the role played by the mental state expressed by (2) in the theoretical and practical inferences from (3), we need, according to Perry, a new sort of content, that is the 'reflexive content'. The reflexive content is a proposition whose truth conditions include, besides the referents themselves, the elements of the representational carrier. Thus, the reflexive content expressed by (2) is the proposition that is true if and only if the person designated by the utterance of 'I' is in Frankfurt. The fact that part of the representation to which the content is ascribed appears in the specification of the content's truth conditions makes the content in question reflexive.

Indexical cases analogous to the knowledge argument are designed to illustrate the thesis that, contrary to what the knowledge argument claims,

---

[2]Perry (2001) uses 'recognitional knowledge' rather than 'indexical knowledge'. To distinguish his position from Loar's however, I will use 'indexical knowledge.'

it is possible to have epistemic progress without expanding the ontology beyond the facts previously learned (physical facts). Perry intends to show that the familiar epistemic gaps in indexical knowledge cases are just like the epistemic gap in phenomenal cases, such as occur in the knowledge argument. Consider Perry's analogous case:

> Gary has been trapped for a month in a windowless hut across from Little America, just off Interstate 80 in western Wyoming. (Little America is a gas station with a restaurant and souvenir shop. It has more gas pumps than any place in the world.) He has memorized an interstate road map. Gary knows all the facts about the locations of things along Interstate 80: the order of states, cities, towns, and villages as one progresses west to east along Interstate 80, from Berkeley through Reno, Salt Lake City, Little America, Cheyenne, Lincoln, and on through the mysterious East. But he isnt allowed to look out of his hut, so he doesnt know where he is. Eventually he escapes. He sees all the gas pumps, realizes he is in Little America, and immediately knows a number of facts that seem to be facts about where things are along Interstate 80 but that he didnt know before. (Perry 2001: 108)

Before being released from his windowless hut, Gary was in a position to assert:

(4) Salt Lake City is southwest from Little America.

After his released, Gary is able to claim:

(5) This place is Little America.

Now he is able to infer:

(6) Salt Lake City is southwest from this place.

If we interpret (4) and (6) by mobilizing only the referential content of the sentence, they would have the same content. But the notion of reflexive content would allow us to properly identity Gary's epistemic progress expressed in (6). The referential content of (4) and (6) is a true proposition

if and only if Salt Lake city is southwest from Little America, whereas the reflexive content of (6) is a true proposition if and only if the designated object by 'this place' is Little America.

Gary acquires a new item of knowledge that he would not be able to deduce from his previous complete objective knowledge. However, the new item of knowledge acquired by Gary is still knowledge about the location of cities and gas stations on the road, that is, the item of knowledge is about the same subject matter of which he already knew everything there was to know. Gary's epistemic progress has distinct practical and theoretical consequences that might be explained by means of the reflexive dimension of Gary's mental states. Mutatis mutandis, according to Perry, this also applies to Mary's case. The distinction between referential and reflexive content would allow the explanation of Mary's epistemic progress upon leaving the black and white room, that is, an explanation in terms of the acquisition of indexical knowledge: not as the apprehension of a new fact but as a new way to apprehend facts she knew before. Perry thinks that physicalism is in danger only when we refuse to recognize a second dimension of meaning, the reflexive content. The premise here is the same used by the two-dimensional semantics and the Fregean classical distinction between sense and reference: a second dimension of meaning is stipulated to deal with epistemic contexts. Applied to Mary's case, Perry argues that the knowledge argument can be reduced to a matter of indexicality. As Perry sees it, if we do not distinguish between two sorts of contents we fail to explain Gary's case as we fail to explain Mary's case. If Mary's new knowledge is a case of indexical knowledge, then it does not require expansion of the ontology beyond physical facts. So the analogy would block the dualist conclusion of the knowledge argument.

Perry's view implies that the concepts mobilized in Mary's new knowledge are indexical or demonstrative. In fact, for Perry, the correct way to express a phenomenal concept is by means of a demonstrative expression, like in the sentence:

(7) *This* is what it is like to see red.

where '*this*' is an indexical, which points to Mary's experience. Gary is like Mary in that he lacks a certain item of knowledge that could only be

acquired in certain contexts like being the center of an egocentric belief.

Perry's answer to (Q1), the question about the conceptual independence of phenomenal concepts, is to say that their independence is explained by their indexicality. The knower needs to place herself in a first-person perspective to make appropriate deductions. This is the case of Gary and Mary. Gary cannot deduce subjective truths like (6) from objective truths like (4) because he does not possess the reflexive concept to grasp (6). Likewise, Mary cannot deduce phenomenal truths from physical truths because she does not possess the phenomenal concept (which, according to Perry, is a reflexive concept after all). Those concepts are different because they have different truth-conditions, the reflexive content of a concept includes the part of the representation to which the content is ascribed, whereas referential concepts fix their referents through descriptions.

Mobilizing demonstrative terms to express phenomenal concepts presupposes that the analogy between phenomenal knowledge and indexical knowledge works. An objection advanced by David Papineau (2007: 113) claims that it is a mistake to associate too closely the demonstrative linguistic expression with the concept expressed by it. In fact, the apparent analogy between (5) and (7), for example, induces the mistake of confusing the concept expressed with its linguistic expression. The concept expressed by the utterance 'this place' is a particular place to which one points demonstratively. In this case, there is harmony between the term used and the concept expressed: The demonstrative term expresses a demonstrative concept. However, the concept expressed by the demonstrative in (7) is not a particular (token) experience, but a *type* of experience. Papineau thinks that demonstrative pointing must be restricted to particular objects. One may point demonstratively to instantiation of a universal, but not directly to the universal. And if we think that a demonstrative expresses a phenomenal concept, we should recognize that there is a mismatch between the concept expressed and the linguistic expression: while the expression is demonstrative, the concept is not.

What is distinctive about demonstratives and indexicals is their context sensitivity, or in Kaplan's terms, the characterlikeness of the term; the referential value of the term will differ in different contexts of use. Phe-

nomenal concepts, contrary to demonstrative concepts, do not seem to be context-dependent in the same way as demonstratives are. Whenever it is exercised, a phenomenal concept refers to the same type of experience. The demonstrative term ('this') in (7) expresses the phenomenal concept $\boldsymbol{Red}_P$ regardless of the context of utterance. That is, when Mary recalls the subjective character of her experience of seeing red she refers to types of experiences instead of tokens of experience. If demonstrative words were used to express phenomenal concepts, we would presumably refer to token concepts. The fact that we might use demonstrative words to express phenomenal concepts (non-demonstrative) does not mean that phenomenal concepts are demonstrative. The reason we do it is, according to Papineau, because 'there is often no publicly established linguistic term to express our concept.' (Papineau 2007: 4).

Of course, 'This experience' may point to released Mary's particular phenomenal state. But what is relevant to the knowledge argument is not the particular experience of Mary, but the type of experience that Mary gains upon leaving the room, the same type of experience that normal human beings undergo when they see a red rose. The demonstrative term in 'this is what it is like to see red' have the sense of this *kind* of experience. If Papineau is right, the correspondent concept is not context-dependent. Papineau's objection can be understood in the form of a dilemma: If the demonstrative term '*this*' in (7) expresses a demonstrative concept, the concept is not of a type of experience; if the concept is the concept of a type of experience, it will not be demonstrative and the use of 'this' should be viewed as a communication proxy, for lack of a better term to designate the concept.

Perry could defend his position by conceding to the negative diagnosis about the demonstrative nature of phenomenal concepts, pointing out that this is irrelevant to the main point. What is important in Perry's account is that the content of phenomenal concepts carries information that is essentially dependent on context (Stalnaker 2008). Perry's main point can then be stated thus: some kinds of information are attached to a context in a way that they can only be acquired by a subject that is the center of this context. This is perfectly compatible with a physicalist ontology.

The contextual dependence of Mary's new knowledge would be immediately assured if phenomenal concepts were demonstrative: the contextual dependence that is distinctive of demonstrative concepts would immediately transfer to phenomenal concepts. Nevertheless, the distinction between reflexive content and referential content initially introduced to give account of indexical knowledge does not have to restrict its use only to demonstrative linguistic expressions. In fact, one could substitute the expression of the phenomenal concept 'this' by a generic term like 'R'. This would concede to Papineau's objection, since R is a general, non-demonstrative concept such that it is possible to ascribe a referential content and a reflexive content to the mental state correspondent to 'R'. Then,

(8) R is what it is like to see red.

is the expression of $\boldsymbol{Red}_P$ but it is not constituted by a demonstrative expression. (8) has a referential content that was apprehended by Mary before she left the black and white room. This knowledge includes the referential information that red objects cause experiences of the type R, since this information can be acquired even by someone who does not have the corresponding experience, as we have seen, this is the content of psychological concepts and not of phenomenal concepts. What is new to Mary is the reflexive content: the singular proposition that is true if and only if the results of her experience with red objects are the type of experience designated by 'R' by the utterer of (8).

There is a contextual dependence character that is not attached to the indexical character of the expression 'R' or the demonstrative character of the concept R. Perry's concession to Papineau is that the term that expresses a phenomenal concept need not have to be indexical nor contain a hidden indexical component. What is important in Mary's new knowledge is its contextual dependence that is assured by the reflexivity that Mary's new knowledge shares with other forms of indexical knowledge.

One objection Chalmers (1999) raises against the indexical account of phenomenal concepts is that there are good reasons to assume that Mary gains more than indexical knowledge upon leaving the black and white room. The first and most important point is that indexical knowledge depends on

the subjective first-person perspective being assumed at the moment of the utterance, but that first-person perspective disappears when we shift to the objective third-person perspective. This, however, is not the case of Mary. Mary's new knowledge does not depend on any perspective. Although she must undergo relevant experiences to acquire certain concepts necessary to make the appropriate deduction, her new knowledge does not disappear if we shift to an objective perspective. The objection starts by pointing out that Gary's indexical ignorance, for example, is not shared by a physically omniscient observer:

> Say that I am physically omniscient but do not know whether I am in the United States or Australia (we can imagine that there are appropriate qualitative twins in both places). Then I am ignorant of the truth of 'I am in Australia,' and discovering that I am in Australia will constitute new knowledge. However, if other people are watching from the third-person point of view and are also physically omniscient, they will have no corresponding ignorance concerning whether I am in Australia. They will know that A is in Australia and that B is in the United States, and that is that. (Chalmers 2004: 186)

The cases involving phenomenal ignorance construed in analogy to the indexical ignorance above behave differently, there is no vanishing of the phenomenal ignorance. Pre-released Mary is ignorant about phenomenal facts, however, a physically omniscient observer might observe Mary and still have the analogous ignorance, he still has no idea what it would be like for Mary to see red. Contrary to the indexical case, phenomenal ignorance does not vanish with perspectival shift. This is strong evidence that phenomenal knowledge is not a sort of indexical knowledge.

The indexical account of phenomenal concepts provides an analysis of Mary's epistemic progress by applying a theory designed to apply to indexicals. However, Perry does not provide a similar analysis of the conceivability argument. Perry's response to conceivability argument does not involve his theory about phenomenal concepts, hence he does not provide an answer to the question about the a posteriority of identity statements.

3.3 The constitutional account

The present section will focus on the version of the phenomenal concept strategy that develops the idea that the special feature of phenomenal concepts is its constitutional character. The constitutional account of phenomenal concepts is formulated by David Papineau (2002, 2007). The constitutional account shares the same initial presuppositions as other accounts of phenomenal concepts. They all hold that there is something special about these concepts and that the epistemic gaps in the anti-physicalist arguments are generated not due to difference in the nature of phenomenal properties and physical properties, but due to the nature of the concepts in terms of which we think about conscious experience.

David Papineau has formulated a theory according to which phenomenal concepts are (at least partly) constituted by the very phenomenal experience to which they refer. For instance, tokens of the phenomenal concept $\boldsymbol{Pain}_P$, which refers to a type of experience, are constituted by tokens of that type of experience. Phenomenal concepts thus refer to the experiences that they exemplify. Papineau's idea is that phenomenal concepts work like quotation marks. The same way that quotation marks carry (mention) the term we want to use, phenomenal concepts use the experience to mention certain experience. So phenomenal concepts can be represented by the structure 'the experience: ... ', where the gap is filled either by the current experience or by an imaginative recreation of the experience.

Papineau thinks that phenomenal concepts work roughly like perceptual concepts. According to his account, we should think of perceptual concepts as *stored sensory templates*. New templates are created when we encounter our referents for the first time. Incoming stimuli form new sensory templates and activates stored templates. If in new encounters we become acquainted with new data about the referent, the sensory template expands to accommodate the new information. When I encounter a woodpecker for the first time, new information will be added to my stored sensory-bird template, as a new template will be created, the woodpecker-template. As I encounter new kind of woodpeckers, new information will be added to my woodpecker-template which is attached to my bird-template. Imagining a bird activates

such sensory templates as well as re-encountering a bird. 'The function of the templates is to accumulate information about the relevant referents, and thereby guide the subject's future interactions with them' (Papineau, 2007: 115).

We use stored sensory templates to think about our own experience. As with perceptual concepts, when we employ a phenomenal concept, a sensory template of the correspondent experience is activated by the experience itself. This activation of templates is what allows us to think about our experience. What activates the template is either a new experience one is currently attending to or a recalling of past experiences either by present experiences or imaginative activity.

How does the constitutional account responds to the anti-physicalist arguments? Standardly, Papineau explains Mary's ignorance in virtue of her lacking the right concepts: Mary cannot deduce phenomenal truths from physical truths because she lacks the phenomenal concept $\boldsymbol{Red}_P$. Mary lacks the relevant concepts because they are experience-dependent, hence they require a stored sensory template to be activated. Requiring the template depends on her visual system having being activated by the perception of the color red. The constitutional character of phenomenal concepts is what responds to (Q1), the question about conceptual isolation of phenomenal concept: because a phenomenal concept is constituted by the experience to which it refers, the phenomenal concepts pick out their referent directly, that is, without any other conceptual mediation. By contrast, physical concepts pick out their reference via other concepts.

Regarding the conceivability argument, the standard response of the phenomenal concept theorists is to break the conceivability and possibility link by claiming that the psychophysical identification (P=Q) or the psychophysical conditional (P→Q) is a posteriori. The physicalist needs to come up with an explanation for the a posteriority of P=Q that challenges a core assumption endorsed by the proponent of the conceivability argument, viz. the concepts involved in the psychophysical identification or conditional are stable. Papineau claims that, although the assumption that a posteriori necessity requires semantical instability holds in standard cases, it does not hold in cases involving phenomenal terms. Phenomenal concepts

are anomalous in this sense. Papineau explains that what is abnormal in phenomenal concepts is their use-mention character: phenomenal concepts use an experience to refer to it:

> Even if phenomenal concepts don't involve direct knowledge of real essences, they will still come out semantically stable, for the simple reason that the use-mention feature lead us to think of the referent as 'build into' the concept itself. Since the concept uses the phenomenal property it mentions, this alone seems to eliminate any conceptual or metaphysical space wherein that concept might have referred to something different. (Papineau 2007: 131)

Papineau's strategy, as well as Loar's, is to deliver an independent reason for the stability of phenomenal concepts: they are stable because they refer directly to their phenomenal properties. This allows phenomenal concepts to be stable even if their possessors are ignorant of the referent's essential features.

In response to (Q2) Papineau says that the semantic stability of phenomenal terms does not require that, once we possess a phenomenal concept, we have a priori access to the essential features of its referent. It requires only that the referent is used in the specification of the concept in order to mention the referent. This response blocks the entailment between conceivability and possibility. If primary intensions associated with phenomenal terms and with physical terms depend on contingent facts about the actual worlds, then the move from the primary possibility of zombies to their secondarily metaphysical possibility will be blocked.

Michael Tye (2009) criticizes two points in Papineau's theory. First, he points out a strange consequence of Papineau's constitutional account: if phenomenal concepts use the experience itself to specify the type of experience, then a consequence would be that the experience is part of the thought. Do we experience pain when we think about pain? This seems like a drastic consequence. When I think about pain, I do not feel the same typical discomfort that I feel when I actually experience pain. This consequence seems to be plausible when we consider visual phenomenal concepts, in fact, when we turn our attention to our experience of a red rose, we invoke the original

experience in our minds—a red image. However, that does not mean in the case of pain that our thought is red, just like it does not mean that out thought hurts. Still, we need to be more fair to the constitutional account, when we claim that experience is part of the concept we do not mean, as Balog writes, "spatial part' but rather part in the sense that it is metaphysically impossible to token the concept without tokening its referent.' (Balog 2012: 25)

The second difficulty pointed out by Tye is that phenomenal concepts refer to types of experience via tokens of experience, that is, the concept picks out a type of experience tokened in an associated replica of the experience. However, a type of experience usually instantiates many tokens of experience. Tye (2009) points to the problem. The experience that is exemplified in a sensory template we use to think about certain conscious episodes is a token experience. The problem is that such a token of my templates for pain exemplifies the general type concept ***Pain***, but also move specific types of pain, like 'sharp pain', 'pulsating pain' etc.. And it also exemplifies other general kinds like phenomenal properties as 'having a phenomenal quality.' For this reason, the experience tokened in the sensory templates cannot play the reference-fixing role, because there are simply too many types of pain that fit a given token. However, Papineau responds to this by claiming that what determines whether a concept will refer to a type or a token template is the experiencer's disposition to attach certain information to the token of that concept. If the subject is disposed to attach particular token of pain, the concept will refer to a particular occurrence of pain, if the subject is disposed to attack kinds of pain, the concept's referent to a kind of pain.

3.3.1 Acquaintance physicalism

Katalin Balog (2012a, 2012b) also defends a version of the constitutional account. Balog believes that what is special about phenomenal concepts is that they afford us, the subjects of experience, 'a special, intimate epistemic access to qualia.' That special relation is understood by Balog as acquaintance. The constitutional account does not provide a physicalistic explanation of acquaintance, but it does deliver a neutral theory regarding

the ontology behind it. The constitutional account, in Balog's view, provides a 'cognitive architecture of mental phenomena that explains acquaintance in a way that it is compatible with physicalism' (Balog 2012b). Balog believes that the constitutional account is the most adequate theory about phenomenal concepts to account for its special features. Balog mentions a number of features, but among them acquaintance is emphasized.

Acquaintance, in the sense of Balog, was first introduced in the literature by Russell (1911). He distinguished between two kinds of knowledge: knowledge by acquaintance and knowledge by description. The latter is the knowledge we acquire, by knowing something via a description. It is to know that some object is the so-and-so. Knowledge by acquaintance, on the other hand, is obtained through a direct cognitive relation between the subject and the object perceived. Objects that are typically grasped through acquaintance are sense-data of a certain object.

These are the two central features of acquaintance: (i) Knowledge by acquaintance is not about something, hence, non-intentional nor representational. It is a form of knowledge that is so direct that for us to grasp it we do not need any conceptual mediation. When we are acquainted with an object we do not form any thoughts or conception of it, we merely establish direct contact with the objects of our perception. Therefore, acquaintance is non-conceptual and non-judgmental; (ii) The object of our acquaintance must exist in order for us to relate to it. In contrast, representation through intentional states does not require the existence of the object represented.

In Russell's view, we might have both knowledge by description and knowledge by acquaintance of the same object. I am acquainted with my neighbors since I know them personally, I have first-hand perceptual knowledge of them, I am aware of them. However, I also know of them that one is an airplane mechanic from Baden and the other one is a saleswoman from Porto. The former pieces of knowledge are by acquaintance, the latter are by description. I can also have knowledge by description of someone without having any acquaintance relation, as in Russell's example, the man who has never met Jack the Ripper but knows that he committed certain crimes in London.

Applied to the phenomenal realm, the acquaintance relation between the subject and her phenomenal states should afford a special access to the phenomenal property, that is, the referent of the phenomenal concept under which these states fall. It is sometimes understood that this relation is so intimate that it reveals to the subject of the experience the nature of the referent.

> If phenomenal concepts are partly constituted by phenomenal states, our knowledge of the presence of these states (in the first person, 'inner' way of thinking of them) is not mediated by something distinct from these states. Rather the state itself serves as its own mode of presentation. (Balog 2012b: 15)

Balog expects the constitutional account to deliver a picture of the cognitive architecture as the token experience that constitutes the phenomenal property, which is the referent of the phenomenal concept of that token experience. Specifically, a phenomenal concept ***Red***, which is constituted by the experience of red itself, refers to the phenomenal property Red. This is supposed to yield a metaphysically neutral framework that is, at least compatible with the hypothesis of acquaintance. Because of the constitutional e immediate character of phenomenal concepts that the Papineau's account delivers, it can serve as cognitive architecture for acquaintance.

An objection has been raised against Balog's version of acquaintance physicalism by Philip Goff (*forthcoming*). Goff argues against physicalism by defending what he calls *the real acquaintance view*. In general terms, the real acquaintance relation is defined by Goff (*forthcoming*) as an epistemically intimate, special relation a subject stands with a property that reveals the essence/real nature of this property.

> The real acquaintance view: If x is acquainted with a property R, then x grasps the real nature of R.

As a result of the acquaintance relation we bear with our phenomenal states, we have immediate awareness of the qualia we are instantiating. The canonical application of phenomenal concepts would yield infallible knowledge (Phenomenal Insight, in Goff's terms). According to the theory proposed

by Balog and Papineau, what explains acquaintance in this sense is the fact that phenomenal concepts are constituted by the very experiences to which they refer. However, this sort of acquaintance physicalist theory cannot accept Goff's thesis that the 'real nature' of phenomenal qualities is disclosed to us by the real acquaintance relation, since the 'real nature' is physical, that is, phenomenal states refer to physical states.

In particular, Goff (*forthcoming*) thinks, that this thesis applies (perhaps uniquely) to phenomenal properties, so that real acquaintance with Q is tantamount to 'grasping the real nature' of Q. Although some physicalists formulate their views in terms of acquaintance, the real acquaintance view is incompatible with physicalism or so Goff claims. For if phenomenal concepts provide real acquaintance with phenomenal properties, we would have grasp of phenomenal properties immediately as physical. But that does not seem to be the case.

Goff arranges the train of thought in the form of an argument against acquaintance physicalism:

(P1) If x is really acquainted with Q, then x grasps the real nature of Q. (RAT)

(P2) The real nature of Q is P (P is a physical property). (Physicalism)

(P3) If A=B and x grasps A, then x grasps B. (Transparency assumption)

(C4) If x grasps the real nature of Q, then x grasps P. (from 2 and 3)

(C5) If x is real acquainted with Q, then X grasps P. (from 1 to 4)

In Goff's original formulation, (P3) and (P4) are rather implicit. Now, Goff claims that (C5) is false. And this means that either (P1), (P2) or (P3) is false, since the argument is obviously valid.

So, in order to keep (P1) and (P3), then (P2) and (C5) must be rejected. But if we analyze the prospects of keeping (P1) and (P2), then we would have to reject (P3) the transparency assumption or accept the conclusion (C5) that we are acquainted with physical properties. If we endorse (C5), contrary to what Goff wants, then the real nature of Q would be revealed to

me as physical. If the real nature of Q is physical (and is revealed as physical) whenever I grasp the property under the concept Q, I grasp it as physical. Grasping the real nature would be indifferent to 'modes of presentation': I would grasp it under the concept Q and also under the concept P. Keeping the transparency assumption is *prima facie* problematic, since the context created by the use of 'grasping' and its cognates is intensional. Thus, if we accept the transparency-assumption, then 'grasping', 'understanding', 'discerning' would have to assume a technical sense. In other words, when we decide to apply the transparency assumption for 'grasping', then we are working with a *de re* understanding of Q. And if we drop the transparency assumption, we would have a *de sensu* understanding of Q (in analogy with *de dicto*). In the *de re* case, we would have to accept (C5) to maintain (P2). However, accepting (C5) dissolves the epistemic gap between phenomenal qualities and physical properties. That is undesirable for the phenomenal concept strategist and for the dualist. In that case, Goff would have its intended reductio.

So, either the physicalist endorses (P5) with a technical (transparent) sense of 'grasp' as in (P3) or rejects (P5) with the ordinary (non-transparent) sense of 'grasp'. In that case (P3) must be false. Goff's argument plays on an equivocation: In order to endorse (P3) he needs a special sense of 'grasping'. In order to reject (P5) he needs the ordinary sense of 'grasping' (on which (P3) is false).

We try now to drop the transparency assumption (P3), for it is only by rejecting transparency that we can deny (C5) and preserve (P2). In that case we may say: being aware that I am instantiating a property Q does not make me grasp what is really going on, if what is really going on is something physical: having the Q experience and attending to it does not reveal to me Q as physical. That is quite unsurprising because, contrary to transparency, normally knowing that a property R is instantiated does not entail knowing that a property S is instantiated, even if R=S. So by denying transparency we open the kind of epistemic gap accepted by both phenomenal concept strategist and dualist. The problem here is that real acquaintance is not normally knowing. As we saw, 'grasp' in the phrase 'grasp of real nature' plausibly has a technical, *de re* sense.

To explain Goff's option for rejecting (C5) and (P2) more has to be said. It would help to say something about the real nature of Q that one grasps by attending to an experience of type Q. What is the 'definitional' essence of pain, that I grasp every time I am aware of my pain? What would have to be true about the real nature of pain so that both (P1) and (P2) are true? What is it for something to be revealed as physical? It cannot be just the property of, for example, stimulation of c-fibers. For in that case, the subject of experience would have a priori, infallible access to the stimulation of c-fibers. Knowledge that your c-fibers are stimulated can only be gained a posteriori. Knowing that my c-fibers are stimulated is not what is required for being acquainted with the phenomenal quality of pain.

So far Goff has produced a valid argument. But its conclusion is a conditional: if we are really acquainted with our phenomenal properties, then they cannot be physical. Now the question is whether we are really acquainted with our phenomenal states.

Katalin Balog (2012a, 2012b) seem to recognize the incompatibility of the notions of real acquaintance (but not acquaintance) and physicalism. She recognizes a clear sense in which the deployment of phenomenal concepts, even in cases of the canonical, first-person application, does not reveal the nature of phenomenal properties, for it does not reveal them *as physical or as functional*. But, she says, in another sense, it does. This sense is expressed by the claim that a token of an experience constitutes the phenomenal concept referring to this (type of) experience. Phenomenal concepts in the constitutional account do not reveal their natures as physical or functional because they do not analyze their referents in physical or functional terms. But in another sense they do reveal the nature of their referent.

> In the canonical, introspective applications of phenomenal concepts, the very phenomenal (i.e., physical or functional) property that is being introspected serves as its own phenomenal mode of presentation. To avoid this equivocation, perhaps it would be better for the physicalist to analyze acquaintance and the substantiality of phenomenal belief in terms of the phenomenal presence of the introspected properties in phenomenal judgments; and not in terms of our direct grasp of the essence of phenomenal properties. This is a characterization of acquaintance that

physicalists and dualists can agree about (Balog 2012: 15).

Of course, Goff would insist here that there is grasp of essence by phenomenal concepts. It is intuitively true that phenomenal concepts put us in direct contact with the phenomenal properties we are instantiating. In fact, it is the fact that phenomenal terms are semantically stable confirms this intuition as Papineau puts it: 'semantic stability goes hand in hand with knowledge of real essences; conversely, if thinkers are ignorant of real essences, they must be using unstable concepts' (Papineau 2007: 131). However, Goff wants to argue that this is incompatible with physicalism. However, in order to his argument to succeed, he must employ a very specific, transparent sense of 'grasping', which is not at all compatible with our use of grasp.

* * *

So far the specific accounts of phenomenal concepts have presented a promising strategy to block the anti-physicalist conclusion of the knowledge argument and the conceivability argument. The general strategy consists in arguing that the *sui generis* character of phenomenal concepts is what explains their conceptual isolation and the a posteriori of identity statements involving phenomenal terms. The *sui generis* character of phenomenal concepts is their perspectival feature, i.e. phenomenal concepts are experience-dependent. I have proposed that specific accounts of phenomenal concepts must respond to two questions:

(Q1) Why are physical and phenomenal concepts independent?

(Q2) How can the psychophysical identification or the conditional be a posteriori if phenomenal terms are semantically stable?

Each assessed account responds to the questions above by means of specific facts about the nature of phenomenal concepts. The recognitional account assigns the ability to recognize and discriminate to phenomenal concepts, which should, according to Loar, assures the conceptual independence and the cognitive isolation of phenomenal concepts. The fact that phenomenal concepts use distinct psychological faculties, according to Hill

and McLaughlin, should respond to (Q2). They provide independent reasons for the cognitive isolation, reasons which are not semantic or modal, so the result is that the psychophysical identification of conditional is a posteriori because they are cognitively isolated, and not because of the way their reference is fixed. This explanation avoids Kripke's obligation of explaining away the appearance of contingence, and Chalmers inference from conceivability to possibility. The indexical account treats phenomenal concepts in analogy to indexical concepts, which are also conceptually isolated from physical concepts. Perry does not provide a response to (Q2), but he does provide a response to (Q1): Mary's case is analogous to many cases of indexical knowledge. Distinguishing between a second dimension of content that grasps indexical content should also be enough to grasp knowledge of release Mary. At last, according to the constitutional account, the cognitive isolation of phenomenal concepts is due to their use-mention character. Because the experience, which is the referent of the phenomenal concept, also fixes its reference, phenomenal concepts are cognitively isolated from physical concepts. We have also considered possible objections to the strategy. And I have concluded that each account can be defended from specific objections. However, the greatest challenge to the phenomenal concept strategy does not concern the specific account of phenomenal concepts, but its general commitments. In particular, the perspectival character is challenged by three powerful arguments against the phenomenal concept strategy, which we will consider on the following chapter.

Chapter 4

General objections

This chapter will examine objections raised against the phenomenal concept strategy. I will argue that they are inconclusive. Therefore, the strategy is a successful response against the anti-physicalist arguments we are considering. I will focus on three general objections: Daniel Stoljar's (2005) objection from a priori and a priori synthesizable conditionals, Derek Ball (2009) and Michael Tye's (2009) argument about social externalism and phenomenal concepts, and the master argument advanced by David Chalmers (2010).

1 Stoljar's objection

Daniel Stoljar (2005) adduces two independent arguments against the phenomenal concept strategy's response to the conceivability argument on the one hand and to the knowledge argument on the other hand. First, against the response to the conceivability argument, Stoljar argues that no special feature of phenomenal concepts can ensure the a posteriority of physicalism. Second, against the response to the knowledge argument, he argues that the thesis that Mary cannot make the appropriate deductions, because she *lacks* the relevant concepts, is false.

Let us consider the psychophysical conditional (PHYS) again:

(PHYS) P→Q [1]

Stoljar characterizes the central project of the phenomenal concept strategy as that of explaining how the a posteriority of the conditional is possible without running into Kripkean difficulties. Stoljar thinks that in order for the phenomenal concept strategy to succeed, the experience thesis or any replacement thesis to the same effect must entail the a posteriority of the psychophysical conditional. Nevertheless, he will argue that the experience thesis does not entail a posteriority of the psychophysical conditional, hence the strategy will fail.

As discussed in the previous chapter, there are facts about the nature of phenomenal concepts which makes them different from physical concepts. Each specific account of phenomenal concepts assigns different properties to phenomenal concepts: the recognitional character on Loar's account, the indexical character on Perry's account, and the constitutional character in Papineau's account. Although not all versions of the phenomenal concept strategy endorse the experience thesis as formulated in the previous chapter, Stoljar's objection is not restricted to the experience thesis. All versions of the strategy are subject to Stoljar's argument, since of all them agree that phenomenal concepts have a perspectival character which makes them different from physical concept. The conceptual difference explains why Q is not entailed a priori by P. For simplicity, though, let us begin by considering Stoljar's argument as applied to the experience thesis.

Stoljar's first objection to be considered applies specifically to the treatment of the conceivability argument by the phenomenal concept strategy. Stoljar argues his point by distinguishing between two ways of understanding the conditional; The *a priori* and the *a priori synthesizable*(Stoljar 2005).

---

[1]Here P is the complete physical description of the world, and Q the phenomenal description of the world. The conditional is meant to express that once every physical aspect of the world is settled, the phenomenal aspects of the world are *necessarily* settled.

> *The a priori*:
> A→B is *a priori* if a sufficiently logically acute person who possessed only the concepts required to understand it, is in a position to know that it is true (478).
>
> *The a priori synthesizable*:
> A→B is *a priori synthesizable* if a sufficiently logically acute person who possessed only the concepts required to understand *its antecedent*, is in a position to know that it is true' (478).

To illustrate this distinction, consider (1):

(1) If y is rectangular, then x has some property or other. (478)

where (1) is clearly *a priori*, but not *a priori synthesizable*, since a logically acute person who knows only the concepts involved in the antecedent of the conditional—the concept ***Rectangular***—is not in a position to a priori synthesize the consequent; A logically acute person who lacks the concept ***Property*** cannot understand the consequent of the condition; hence, she is not in a position to understand the conditional. Being a priori synthesizable entails being a priori, but not the other way round, a priori conditionals may fail to be a priori synthesizable as in (1).

To avoid the conceivability argument, the phenomenal concept strategist must hold that the conditional (PHYS) is not a priori. The crucial premise of the phenomenal concept strategy is that the experience thesis entails that (PHYS) is not a priori. This is because, for physicalist to be true, P must entail Q. Since we concede to the epistemic gap of the anti-physicalist arguments, the entailment is not a priori, the only option left is to show that the conditional is a posteriori. The experience thesis is crucial because is marks the cognitive difference between phenomenal concepts and physical concepts, as we have seen in the previous chapter. So, the project of the phenomenal concept strategy is to argue that because phenomenal concepts are experience-dependent, they are conceptually independent from physical concepts, hence the conditional is a posteriori. Stoljar concedes that being a posteriori is equivalent to be not a priori. His argument is: the phenomenal

concept strategy claims that the experience thesis entails that PHYS is not a priori. But then he shows that the experience thesis entails only that PHYS is not a priori synthesizable and this is irrelevant for the a posteriority of PHYS. It is irrelevant because one can find many examples of a priori propositions like (1) which are not a priori synthesizable and still a priori.

His conclusion is that all that the experience thesis entails is that (PHYS) is not a priori synthesizable. Granting that (PHYS) is not a priori synthesizable does not eliminate the possibility that (PHYS) is not a priori, like in (1). 'So there seems to be a logical gap in the suggestion that the experience thesis tells us that the conditional is a posteriori. What we wanted was a reason to suppose that it is was not a priori. What we have is a reason to suppose that it is not a priori synthesizable' (479). When failing to point out that (PHYS) is *not* a priori, the experience thesis does not provide an answer to the conceivability argument.

Proponents of the strategy could object to this line of reasoning by saying that examples like (1) are relevantly different from (PHYS), because (1) does not involve phenomenal concepts, while (PHYS) connects the antecedent which contains only ordinary concepts and the consequent which contains phenomenal concepts. Thus, the two ways to interpret the conditional could be ignored. But Stoljar has a follow-up argument to the same conclusion. Stoljar asks us to consider (2), a statement which is relevantly like the conditional in that it connects an antecedent containing an ordinary concept with a consequent containing a phenomenal concept (possession of the concept that appears in the consequent of the conditional requires experience):

(2) If x is a number then x is not a red sensation. (479)

Sentence (2) is, like (1), clearly a priori, but not a priori synthesizable. Someone who lacks the concept ***red sensation*** required to understand the consequent of the conditional cannot deduce the consequent from the antecedent. The problem for the physicalist is if PHYS turns out to be a case like (2): clearly a priori, but not a priori synthesizable. Then, the physicalist would have failed to show the a posteriority of PHYS through the experience thesis.

The lack of a priori synthesizability would explain the conceptual independence between phenomenal concepts and physical concepts. But, in this case, conceptual independence does not lead to lack of a priori connections. The phenomenal concept theorist could consider that the distinction works in her favor, at least in the case of the knowledge argument: Mary's inability to deduce Q from P is not explained by claiming that Q and P refer to different properties, but it is explained by the fact that Mary simply cannot synthesize those truths a priori. She cannot a priori synthesize Q from P because she lacks some crucial concepts to understand the consequent of the conditional (a phenomenal concept). However, Stoljar advances an independent objection against the strategy's treatment of the knowledge argument (which will be addressed at the end of this section). For now, the distinction above is aimed at the strategy's response to the conceivability argument.

The proponent of the phenomenal concept strategy could argue that Stoljar's objection works only against versions of the strategy that are committed to the experience thesis. Do other versions also fail in view of Stoljar's distinction between the a priori and the a priori synthesizable? Stoljar analyses a specific version of the strategy to show that not only the experience thesis, but any replacement thesis fail in view of the distinction. In the following I shall consider a version of the theory advocated by Hill and McLaughlin (1999)

According to Hill and McLaughlin (1999), phenomenal concepts and physical concepts are governed by different epistemic constraints and presuppose use of different faculties. One difference between phenomenal concepts and physical concepts is that the former are *self-presented*:

> *Self-presentation thesis*:
> It is a conceptual truth that if I have a red sensation, and if I have the concepts and focus my attention on the matter, I will thereby come to know that I am having. (Stoljar 2005: 483)

On the other hand, 'it is not a conceptual truth that if I am in some overall *physical* conditional P, and if I have the concepts and focus my attention on the matter, I will thereby come to know that I am in P.' (Stoljar 2005: 483) This is to say that, if there is a conditional, whose antecedent contains an

ordinary concept and the consequent contains a self-presenting concept, then the conditional cannot be a priori. Stoljar grants that phenomenal concepts might be self-presenting in the above sense. Still, he holds that the self-presentation thesis, like the experience thesis, cannot explain the posteriority of the psychophysical conditional. Stoljar claims that applying the negation of a phenomenal concept ***not a red sensation*** requires possession of the phenomenal concept ***red sensation***, hence the possession of negations of phenomenal concepts are experience-dependent, or, according to the Hill and McLaughlin's version, the negation of concepts of experience requires the possession of self-presenting concepts. So ***not a red sensation*** is self-presenting just as ***red sensation*** is. On the other hand, possession of the concept ***number*** is clearly not self-presenting (nor experience-dependent). The sentence (2) contains theoretical concepts on the antecedent and self-presenting concepts in the consequent just like in P→Q. If the reason that P→Q is a posteriori and appears to be contingent in Hill and McLaughlin's version of the strategy is that these two different kinds of concepts are entailed, then (2) should also be a posteriori and have an appearance of contingency. Nevertheless, (2) is clearly a priori, so the self-presenting thesis cannot be correct nor can the experience thesis.

The physicalist can respond to Stoljar's observations by claiming that cases like (2) are clearly a priori because the negation of a phenomenal concept, such as ***not red sensation*** does not require possession of the phenomenal concept ***red sensation***. Mary, who lacks all kinds of color related phenomenal concepts, including ***red sensation***, is in a position to know a priori that (2) is true if she possesses at least partial understanding of the consequent, that is, if she possesses the concept ***Sensation***. She knows that, if something is a number, then it is *definitely* not a sensation of any kind. She may also possess those concepts second-handedly (through testimony). She may know that sensations are usually associated with perceptual states, while numbers are not, and that should be enough for Mary to know (2) a priori. Stoljar's reasoning does not pay due attention to the disanalogy between (PHYS) and (2). This may be brought out by considering analogous strategies concerning other kinds of concepts. Consider, for example, a natural kind concept strategy applied to the following conditional involving

a natural kind term.

(3) If x is H_2O then x is water.

The sentence (3) is necessary and a posteriori because it connects a theoretical concept (***H_2O***) to a natural kind concept (***water***). But now consider:

(4) If x is a number then x is not water.

(4) connects a theoretical concept to the negation of a natural kind concept, but (4) is clearly a priori. Although the negation of a natural kind is not a natural kind, the question is whether the negation of a natural kind concept requires possession of a natural kind concept. According to Stoljar's thesis, the natural kind concept ***water*** is required to understand the consequent of (3) but not the antecedent just as it is with (4). But (4) is a priori like (2), and thus, if we were to follow Stoljar's reasoning, we would conclude that the natural kind concept strategy fails. There is a clear disanalogy between (PHYS) and (2) as between (3) and (4).

Another disanalogy between cases like (PHYS) and other conditionals which connect consequents containing experience-dependent concepts and antecedent containing only ordinary concepts is (5).

(5) if x is a square circle then x is a red sensation.

Because the antecedent is a priori false, the whole sentence is a priori false. There is no need to understand the consequent in this case. This illustrates another type of sentences structurally like (PHYS) but with a different epistemic status. The fact that (4) and (5) are a priori does not undermine the posteriority of PHYS, it shows only that we do not use the same criteria to evaluate the epistemic status of PHYS that we are to evaluate (4) or (5).

In sum, in order to respond to the conceivability argument, the proponent of the strategy needs to argue for the a posteriority of (PHYS). The phenomenal concept strategy holds that the experience thesis or any replacement thesis drawing on the perspectival nature of phenomenal concepts is the key feature to deliver the a posteriority of the psychophysical conditional. The distinction between a priori statements and a priori synthesizable statements is introduced by Stoljar in order to show that the best

that the experience thesis accomplishes is to show that physicalism is not a priori synthesizable. The latter is admittedly irrelevant to the explanation of why it is conceivable that the psychophysical conditional is contingent, without being possible. However, there is a clear disanalogy between (PHYS) and other cases involving ordinary concepts. If Stoljar's argument worked against the phenomenal concept strategy, it would also generate undesirable results for the 'natural kind concept strategy'. The core problem with his argument is to assume that negations of phenomenal concepts require possession of that concept.

The second part of Stoljar's argument concerns the treatment of the knowledge argument by the proponents of the strategy. Stoljar thinks that the knowledge argument can be reformulated such that the *new concept explanation* is not a satisfactory answer to the problem. Versions of the phenomenal concept strategy that are committed to the experience thesis explain Mary's epistemic progress in terms of the acquisition of a new concept. Only after leaving the room can Mary acquire the new concept that enables her to make the appropriate deductions. At this point, the distinction introduced by Stoljar to work against the strategy's response to the conceivability argument could work in favor of the strategy's treatment of the knowledge argument. The experience thesis explains why phenomenal truths are not a priori synthesizable from physical truths, and that would be enough to explain the epistemic gap between phenomenal concepts and physical concepts. However, Stoljar rejects this line of reasoning arguing that there is an independent reason to reject the phenomenal concept theorist's explanation of Mary's ignorance. Stoljar offers a different thought experiment in order to show that even if Mary possesses the relevant phenomenal concepts, she would still not be able to deduce phenomenal truths from physical truths. Since the phenomenal concept strategy's treatment of the knowledge argument is to explain Mary's ignorance in terms of her lacking the phenomenal concept, the strategy would fail.

Stoljar asks us to consider *experienced Mary*. She is just like Mary for the first part of the story; after experienced Mary is released from the black and white room, she has color experiences and, because of that, she is able to apply the relevant phenomenal concepts. She is, later, recaptured and re-

turned to her room. After she returns to her room, experienced Mary suffers a process of selective amnesia: she forgets the correct application of phenomenal concepts. She still knows what it is like to see green, thus she still possesses the phenomenal concepts that she acquired during her short period of freedom. However, she fails to make associations like 'looking at Granny Smith apples typically causes green sensations' or 'having arthritis causes pain'. Experienced Mary cannot deduce phenomenal truths from physical truths *even though she possesses the corresponding phenomenal concepts.* In Stoljar's view, she knows the antecedent of the conditional (PHYS) but not its consequent. Stoljar wants to show that this new scenario turns the acquisition of phenomenal concepts irrelevant to explain Mary's ignorance.

I think that the tale of experienced Mary undermined the fundamental premise of the knowledge argument against physicalism, viz. that Mary has complete physical knowledge. What seems to be missing in this version of Mary is information which belongs to the antecedent of the conditional P→Q, not to the consequent. Deleting part of her memory turns the physical knowledge of experienced Mary incomplete, hence, experienced Mary would not possess the relevant concepts to understand the antecedent of the conditional. Without means to understand the antecedent, Mary cannot understand the conditional. It is safe to say that the correct application of phenomenal concepts would be available to experienced Mary, since it is, in a sense, information belonging to the objective, physical domain. Mary must not undergo any experience to know that a deep cut in one's skin *typically* causes pain or that looking at granny smith apples *typically* causes green sensation, and that looking at red fire hydrant *typically* causes red sensations. Original Mary possesses all this knowledge inside the room, which still does not enable her to deduce phenomenal truths from physical truths.

If we add new phenomenal concepts to Mary's set of beliefs, but later we discard her beliefs about the application of those concepts, we discard knowledge that is crucial for Mary to understand the antecedent of the conditional not the consequent, knowledge that used to be part of her complete physical description about the world. Experienced Mary gains phenomenal concepts, but loses ordinary concepts. Experience would only enable Mary to *master* the concepts she possesses partially while still inside the

room. However, if experienced Mary already possessed such concepts, the case Stoljar presents to us is not a case in which she knows the antecedent of the conditional but cannot deduce the consequent. It is a case in which information belonging to the antecedent is omitted. Mary cannot deduce phenomenal knowledge from incomplete physical knowledge. At this point one could ask why should the examples of correct phenomenal belief application should be considered physical knowledge given that ***pain*** is a phenomenal concept. Now I should remind us of the distinction made at the beginning of chapter two between phenomenal concepts and psychological concepts. Not all concepts about phenomenal states are phenomenal concepts. One can have non-phenomenal concepts about phenomenal states, like color blinded people believe that granny smith apples are green, without ever having experienced colors.

In sum, I argued against Stoljar by undermining his argument from experienced Mary and by showing that Stoljar's strategy to deflate the phenomenal concept strategy would also undermine many other strategies, among them, the widely accepted strategy to explain the a posteriority of theoretical identity statements.

2 Ball's argument against phenomenal concepts

We will now address another general and independent objection to the phenomenal concept strategy posed by Michael Tye (2009) and Derek Ball (2009). Ball and Tye argue that phenomenal concepts are not special vis-à-vis physical concepts. For that reason, they will argue that there are no phenomenal concepts *sui generis*. In an attempt to deflate the strategy, they attack its central feature: the experience condition or some replacement of it designed to serve the same purpose. The reasoning is straightforward: According to phenomenal concept strategists, what makes phenomenal concepts 'special' is that they are experience-dependent or perspectival in a way that other concepts are not. If it is possible to possess a phenomenal concept without undergoing the corresponding experience, then the so-called phenomenal concepts are not distinct from physical concepts; and the strategy fails. Ball and Tye claim that phenomenal concepts may have *deferential*

possession conditions. They think that *social externalism* can be applied to phenomenal concepts and that alone would make them as ordinary as other concepts leading to the drastic conclusion that *there are no phenomenal concepts* as a separate category.

2.1 Concept possession and social externalism

Tyler Burge (1979) makes use of some compelling examples to show that social institutions play a central role in determining the contents of our thoughts, including those that do not involve natural kind concepts. Burge's examples lead to the conclusion that the content of our beliefs are partially determined by our linguistic environment and for that reason, we can say that we possess a certain concept even if we have a poor conception of that concept.

Let us consider Tyler Burge's original example. Bert complains to his doctor about having arthritis in his thighs. Bert does not know that arthritis is the inflammation that occurs only in the joints, he is not aware that one cannot have arthritis in thighs. When Bert utters:

(6) I have arthritis in my thigh.

he expresses a false belief. The doctor *corrects* Bert by saying that he does not have arthritis in his thighs, Bert immediately accepts his doctor's correction and corrects his *misconception* of ***arthritis***. It is plausible to claim, Burge argues, that Bert shares the concept ***arthritis*** with his doctor by deferring to the doctor regarding the concept's extension. Bert possesses the concept ***arthritis*** even though he has a poor conception of that concept. He possesses the concept in virtue of his interaction with his linguistic community. The thought experiment is designed to argue in favor of the social externalism thesis, which claims that it is possible to possess a concept in virtue of our social interactions with our linguistic community and not solely in virtue of our intrinsic properties.

An alternative explanation of Burge's example, anticipated by Burge himself, is to say that Bert did not express a false belief in the first place since he did not share the concept ***arthritis*** with his doctor. Rather, his uttering of 'arthritis' expressed a different concept: ***t-arthritis*** (twin-arthritis)

where ***t-arthritis*** can occur both in the joints and in the muscles (or perhaps, only in the latter). What Bert learns from his doctor is a new concept: a concept, whose referent is the inflammation that occurs only in the joints, not in muscles. The alternative hypothesis is that there are two concepts involved in Bert's conversation with his doctor: ***arthritis*** and ***t-arthritis***. The concept of the doctor's beliefs is ***arthritis***, while the concept of Bert's beliefs is ***t-arthritis***. Burge renders this alternative to be highly implausible. There is strong evidence against the claim that Bert and his doctor entertain different concepts. They share the concept ***arthritis***, even though they share it under different degrees of mastery. The plausibility of social externalism and the failure of the alternative dual-concept hypothesis follows, on reflection, from our everyday use of concepts.

First, there is a general agreement that concepts are *public entities*. We can communicate our thoughts and understand what other people expect of us because we share the concepts we deploy. If concepts were private, it would be quite difficult for people to share thoughts. Second, evidence for the publicity of concepts is the fact that we are able to agree and disagree in conversations. The doctor disagrees with Bert's auto-diagnosis that he has arthritis in his tights. This points to the fact that they share the concept ***arthritis***, and this rules out the possibility of both Bert and his doctor applying different concepts in conversation, this talking passed each other. Genuine agreement or disagreement requires shared concepts. If they share the concept ***arthritis***, one of them possesses an impoverished conception of that concept, and the other, the expert, masters the concept in question. Although Bert has an impoverished conception of the concept, we may still say that he possesses the concept—though he has only a partial understanding of it. What we may not say is that he lacks the concept ***arthritis*** altogether. If that were the case, Bert would not be corrected by what the doctor says. If they did possess different concepts, Bert would be able to reject the doctor's correction by saying that the doctor is playing with the meanings of words. Burge's evidence for the thesis that we share concept, even under different mastery conditions is that (i) concepts are expressible in a public language; that (ii) we can agree and disagree about the content of our thoughts; and that (iii) we allow ourselves to be corrected in

cases of disagreement i.e., concepts are over-(under)extended.

Concepts that can be possessed in virtue of such interactions with the linguistic community are often called *deferential*. One *defers* to experts regarding the concept's extension (i.e. experts use the concepts non-deferentially[2] as Bert defers to his doctor regarding the extension of ***arthritis*** and as we defer with respect to many, if not most of our everyday concepts. There are different degrees of concept possession. A school teacher in physics possesses the concept of an electron to a higher degree than her pupils, while she in turn defers for a higher degree of mastery to researches specialized in nuclear physics. Social externalism claims that concept possession *does not* require full mastery of concepts. We may acquire concept simply by linguistic deference.

Now consider Bert's example transposed to color concepts: A blind person may exercise the color concept ***red*** in his thoughts by, for example, ascribing color-experiences to someone who is not blind. Of course, his conception of the phenomenal concept $\boldsymbol{red}_P$ is impoverished. But to possess an impoverished conception of ***red*** does not mean that one lacks ***red*** altogether. It only means that the subject *defers* to experts regarding the extension of the terms used. As it happens, we do defer with respect to most of our ordinary concepts. The general idea is that I can possess some concept C even if my conception of C is impoverished. Ball argues that it is plausible to apply Burge-style considerations to phenomenal concepts, which would make phenomenal concepts too deferential. But then the experience thesis is threatened. If social externalism is true of phenomenal concepts, then pre-release Mary may possess the relevant concept in a deferential manner while still inside the black and white room. But if Mary can possess phenomenal concepts deferentially, then they are not experience-dependent, hence they hold no special feature vis-à-vis ordinary concepts.

So Ball's argument can be formulated in the following way:

---

[2]There is some dispute over the fact that non-deferential uses are easy to spot or if there is deferentiality ad infinitum.

(P1) Experience is a necessary condition for the possession of phenomenal concepts.

(P2) Phenomenal concepts can be deferentially acquired.

(P3) If phenomenal concepts may be possessed deferentially, then it is not required that one undergoes certain experience in order to possess a concept of experience.

So given (P2) and (P3), the experience thesis (P1) is false. Ball also assumes that:

(P4) If experience is not a necessary condition for the possession of phenomenal concepts, then there is no special feature that distinguishes phenomenal concepts from physical concepts.

Thus he concludes that:

(C5) There are no phenomenal concepts as distinct from physical concepts, whence the phenomenal concept strategy fails.

Ball's central claim is that the phenomenal concept theorist is faced with a dilemma: either she rejects externalism about content in general, or she admits that there is no special category of phenomenal concepts. This line of reasoning appeals to the public character of concepts. If concepts are public, then they seem at least *prima facie* incompatible with such an internalist (experiential) individuation account.

There are a few ways to deny one of the premises above. If we want to keep experience-dependence as a phenomenal concept constraint, i.e., (P1), then we have to drop (P2) the assumption that phenomenal concepts are deferential, which means that one may possess the concept C even if one has a deeply impoverished conception of C. Another option is to drop the first premise that states that experience is a necessary condition for *possession* of phenomenal concepts. In this case the advocate of the phenomenal concept strategy needs to present a different property of phenomenal concepts by which they are distinguished from other concepts. In what follows,

I would like first to explore the prospect of denying that phenomenal concepts are deferential without giving up on social externalism. Then, I want to examine a suggestion of weakening the experience thesis so that it accommodates social externalism. My conclusion will be that Ball's argument is twice flawed. First, Burge-style arguments do not show that phenomenal concepts are deferential; there is, in this respect, a deep disanalogy between phenomenal concepts and physical concepts. Secondly, even if we grant that they are deferential, a slight modification of the experience thesis will suffice to block the drastic conclusion that phenomenal concepts do not exist.

Let us start with Ball's reasons for the application of social externalism to phenomenal concepts. This thesis is based on a scenario in which Mary possesses the relevant phenomenal concepts still inside the room. She acquires these concepts in virtue of interactions with normal perceivers through her computer, the lectures she watches, the books she reads. This is made possible by a conjunction of assumptions. First, Ball (Ball 2009) adapts a very general conception of *concept possession*: A subject possesses a concept if she is able to exercise that concept in her thoughts. Secondly, it is argued that phenomenal concepts are expressible in a public language. When Mary leaves her room she can entertain certain beliefs expressed e.g. by (Ball 2009: 946):

(7) That is what it is like to see red.

Let us say that the term 'red' in (7) expresses $\boldsymbol{Red}_P$.[3] Now, consider some other sentences (Ball 2009: 947):

(8) Ripe tomatoes typically cause experiences of red.

(9) What it's like to see red resembles what it's like to see black more than it resembles what it's like to hear a trumpet playing middle C.

---

[3]'That' in (7) designates the experience of red. However, it is at least controversial that the demonstrative expression 'that' refers to the phenomenal concept. 'That' cannot express an indexical concept, for Mary could have learned indexical truths while inside the room. For this reason, we shall assume that 'red' in (7) and in the following sentences express the phenomenal concept $\boldsymbol{Red}_P$.

(10) If x is a number then x is not what it's like to see red.

The occurrences of 'red' in (8)-(10) may also be viewed as expressing the phenomenal concept $\boldsymbol{Red}_P$ when uttered by normal perceivers. Nevertheless, it is plausible to say that pre-release Mary could also entertain beliefs expressed by (8)-(10). However, according to the experience thesis, if Mary does not undergo the corresponding experience, she could not possess phenomenal concepts expressed by 'red' in (8)-(10). The phenomenal concept proponent may say that in Mary's mouth 'red' as in (8) - (10) expresses a different, non-phenomenal concept ***Red***. However, this strategy is the same as the one discussed and dismissed by Burge in the arthritis case: the idea that Bert and his doctor express different concept when they utter 'arthritis'. One problem Ball points out is the counter-intuitive interpretation of apparently contradictory utterances made by Mary:

(11) I do not know what it is like to see red. (uttered before her release.)

(12) I know what it is like to see red. (uttered after her release.)

According to the interpretation that ascribes different concepts pre-release Mary and release Mary, (11) and (12) would express the following thoughts:

(11′) I do not know what it is like to see ***Red***.

(12′) I know what it is like to see $\boldsymbol{Red}_P$.

According to Ball's view of the phenomenal concept strategy, if Mary does not possess $\boldsymbol{Red}_P$ inside the room, then the concepts expressed by 'red' in (11) and (12) are different. 'Red' in (11) expresses the non-phenomenal concept *Red*, whilst in (12) it expresses $\boldsymbol{Red}_P$. This would produce the result that (11) is *not* the negation of (12). Ball thinks that this is highly implausible.

Suppose now that pre-release Mary utters (11) and some of her friends (who have seen colors) utter (12). Another problem is that, if (11) is not the negation of (12), because 'red' expresses different concepts in each utterance, the utterance of (11) by Mary and the utterance of (12) by her friends would not express a cognitive difference between Mary and her friends. To avoid

this problem, the phenomenal concept theorist could adopt the dual-concept strategy that associates two concepts of the term 'red' in (12). Now they are able to disagree because 'red' in (12') would express both concepts, ***Red*** and $\boldsymbol{Red}_P$. However, as Ball points out: 'introspection renders this claim implausible. Mary is not aware of having two sets of thoughts' (Ball 2009, 953).

Ball's arguments are supposed to show that phenomenal concepts are deferential like ordinary concepts. Those two arguments: (i) contradictory sentences and (ii) introspection are arguments analogous to that mobilized by Burge to argue for the deferentiality of concepts. If they can be correctly applied to concepts of experience, they should, according to Ball, lead to the conclusion that phenomenal concept are like ordinary concepts.

Against the counterintuitive hypothesis that (11) and (12) do not express contradictory thoughts, phenomenal concept theorist could say that, in fact, those statements do not express contradictory thoughts and that this result, although counterintuitive, is not as implausible as it seems. Ball anticipates such a proposal. He considers that the phenomenal concept theorist could argue that the apparent contradictory sentences (11) and (12) are like (13) and (14).

(13) I do not know that Hesperus is bright.

(14) I know that Hesperus is bright.

in the Fregean case above, 'Hesperus' in (13) and in (14) might express different concepts, hence thoughts expressed by (13) and (14) would not be contradictory. Consider someone who has only the concept ***Phosphorus*** but not ***Hesperus***. In this context, the sentences would be better expressed in:

(13′) I do not know that ***Hesperus*** is bright.

(14′) I know that ***Phosphorus*** is bright.

Against his analogy, Ball adduces the disanalogy between Fregean cases like (13) and (14) and phenomenal cases like (11) and (12). Suppose that, upon leaving the room, Mary says:

(15) I used to wonder what it is like to see red, but now I know.

(15) does not strike us as strange. In a sense, there is denotation to two different concepts. In (15) 'red' denotes ***Red***, for Mary used to have only the concept *Red* in her room. The ellipsis after 'I know' hides another occurrence of 'red': *but now I know what it is like to see red* in which 'red' denotes $\boldsymbol{Red}_P$. But when considering the analogous Fregean case

(16) I use to wonder whether Phosphorus was bright, but now I know.

it is not clear anymore how someone can utter (16) truthfully, since someone who did not know that Phosphorus was Hesperus could not utter (16). The ellipsis after 'I know' here eliminates the possibility of 'Phosphorus' referring to ***Hesperus*** since the utterer of (16) did not know that Phosphorus is Hesperus, so she could not wonder whether Phosphorus was bright, making the first part of the sentence false.

Veillet (2012) explains this by noticing that it is a mistake to compare ***Red*** and $\boldsymbol{Red}_P$ to ***Hesperus*** and ***Phosphorus***. Our intuitions about the truth of (15) and (16) diverge because the thoughts expressed by them are radically different.

> [The utterer of (16)] is singling out a particular fact about Phosphorus that she used to wonder about and that she now knows-the fact that Phosphorus is bright. In uttering [(15)], Mary is not singling out any one particular fact about r [phenomenal property] that she used to wonder about and that she now knows: she is not saying that she used to wonder whether some thing was true of r; she is not saying that she now knows that fact to be true. Instead, she is saying that she used to wonder, much more generally, what R*[$\boldsymbol{Red}_P$] feels like, and what she now knows is, generally speaking, what R* feels like. However, wondering about or knowing what R* feels like is interestingly unlike wondering whether or knowing that, e.g., seeing red feels like R*. (Veillet 2012: 117)

If phenomenal cases were analogous to the Fregean cases, we could say that Mary could not think (15) for the same reasons: If in (15) 'red' expresses

Red$_P$, Mary could not wonder what it is like to see ***Red***$_P$, for Mary did not possess this concept.

Ball and Tye consider yet another argument for the plausibility of the thesis that phenomenal concepts are deferential: someone may think that she is in pain, when, in fact, she has only a feeling of pressure or even nausea.

> Similar arguments can be developed as regards other candidate phenomenal concepts. For example, consider the concept pain'. It is possible to possess this concept despite having an inaccurate conception of PAIN. For example, the belief that nausea is a type of pain is surprisingly widespread. Those subject to this belief can agree and disagree with others, and in many cases would be willing to change their views in response to correction. We attribute to them beliefs using the word 'pain', and it is very plausible that we should regard them as possessing the PAIN, the very same concept as normal speakers and scientific experts. (Ball 2009: 954)

The idea here is that the fact that the application of phenomenal concepts can be over- or underextended is what supports deferentiality of phenomenal concepts. But this claim is faced with a difficulty when dealing with this claim. The difficulty concerns the possibility of over- or extending phenomenal concepts. It is at least controversial that social externalism may be applied to phenomenal concepts. First, we have the strong intuition that being in pain is an incorrigible mental state. We have full authority of knowing in which state we are. If so, then it is highly implausible that some expert on pain, say, a doctor, might be able to correct us on that matter. Second, I would like to consider a more technical question concerning, how the reference of those concepts are fixed. Concepts of conscious experience, such as our concept of pain, refer to its extensions not via contingent properties of 'pain' but via its essential properties. This is in contrast to physical concepts, such as water, which are fixed via its contingent properties. Because they are fixed via contingent properties, we can be mistaken about their reference. Fixing reference via contingent properties opens the possibility of misconceptions of such concepts. This is to say, as explained in Chapter Two, that phenomenal concepts are semantically stable. Because 'pain' is a semantically stable term, there is no possibility of being mistaken about or

having an impoverished conception of its reference. After all, what would be an impoverished conception that allows us to poorly deploy the concept ***Pain***?

The concept referred to by Ball in the borderline cases between pressure and pain, and in Mary's ascription of color-beliefs to people outside the room may be considered a concept which refers to phenomenal states, but it is not a phenomenal concept. Although these concepts refer to phenomenal states, they are not formed from a first-person perspective. The fact that the concepts merely refer to phenomenal states does not make them phenomenal concepts. They must be experience-dependent and be formed from a first person perspective. In fact, phenomenal concept theorists admit the possibility of acquiring non-perspectival concepts about phenomenal states. They admit that Mary can truthfully utter sentences like

(8) Ripe tomatoes typically cause experiences of red.

(9) What it's like to see red resembles what it's like to see black more than it resembles what it's like to hear a trumpet playing middle C.

without possessing any phenomenal concepts. Mary has a non-phenomenal concept about a phenomenal state of a third party. Just like it is possible for a blind person to possess non-phenomenal concepts about phenomenal states of other people.

I have argued that there is no reason to accept (P2), i.e. the thesis that phenomenal concepts can be possessed deferentially. The arguments mobilized by Burge for the deferentiality of ordinary concepts fail when applied to the phenomenal case, because of the disanalogy between phenomenal and non-phenomenal cases. This does not mean that we do not accept social externalism, it just means that it is not true of phenomenal concepts.

2.2 The concept-mastery objection

Even if the physicalist grants (P2), i.e. the thesis that phenomenal concepts can be possessed deferentially, there is still the possibility to block Ball's conclusion by changing (P1), i.e. the experience-dependence constraint, thereby changing the commitments of the phenomenal concept strategy.

This defense of the phenomenal concept strategy vis-à -vis Ball and Tye's objection is proposed by Torin Alter (2013). Alter grants (P2) that social externalism can be applied to phenomenal concepts, for the sake of the argument, and raises a further issue. He suggests that the phenomenal concept constraint can be reformulated to avoid Ball's objection from social externalism. The experience thesis should be formulated not as a condition for concept possession, but as a condition for concept mastery. This is called the *concept-mastery objection*. The modification in the constraints for phenomenal concepts is that 'it is mastery, not mere deferential possession of phenomenal concepts that normally requires having the relevant experiences.' (Alter 2013: 486). The concept-mastery objection delivers an alternative explanation for Mary's epistemic progress: after Mary leaves the room, she comes to *master* concepts she already possessed inside the room. Inside the room she had an impoverished conception of red, now, after her released, she enriches her conception; she is now a full master of color concepts. Thus, Alter distinguishes between two kinds of knowledge:

> *Knowledge*$_M$: Knowledge under concepts that the knower possesses with mastery (non-deferentially).
>
> *Knowledge*$_P$: Knowledge under concepts that the knower possesses with or without mastery (deferentially or with partial understanding).

Knowledge$_M$ corresponds to the knowledge of experts, whereas knowledge$_P$ corresponds to the knowledge of laymen. The distinction is not exclusive of concepts of experiences, on the contrary, it is a general distinction that should be applied to all concepts. For example, the student uses the concept ***Electron*** deferentially, whereas his physics teacher uses the concept to a higher degree, while she in turn defers for a higher degree of mastery to researchers. [4]. The student has a poor understanding of what an electron is.

[4]It is not so clear whether there is the possibility of a non-deferential use of concepts, since it is always possible to correct or expand our knowledge about some subject matter. The idea of non-deferential application of concepts suggests that there is a point in which

She may know that it is a subatomic particle with negative charge. However, she really does not understand what this actually means. She usually defers the application of that concept to experts: by 'electron' she mean whatever the physicists mean by it. In contrast, a physicist need not to defer to anyone about her use of ***Electron*** since she has, presumably, an extensive knowledge of ***Electron***. She knows$_M$ (in this strong non-deferential sense) the mass of an election, what being a subatomic particle means etc.. Alter suggests that the phenomenal concept strategist ought to abandon the explanation of Mary's epistemic progress in terms of acquisition of a new concept and replace it with the concept-mastery explanation: When Mary leaves the room, she does not gain a new concept, for according to social externalism, she already possessed the relevant concept inside the room, even though her conception of it was impoverished. The novelty is that now she has enhanced her conception of ***Red***$_P$, now she possesses the concept non-deferentially, like people outside the room, that is, like the experts on phenomenal red (people who have seen red). Moreover, Alter thinks that the epistemic progress claimed in the knowledge argument needs to be one of mastery. Only in this was may the knowledge argument have metaphysical bite:

> Suppose that Mary's epistemic progress were construed in terms of her gaining *mere* knowledge$_P$, that is, in terms of her acquiring rather than mastering phenomenal color concepts. In that case, her progress would fail to provide even prima facie grounds for inferring non-deducibility or any epistemic claim that might plausibly entail strong metaphysical conclusions such as non-necessitation. Instead, it would be clear that a psychological explanation is called for. The inference from epistemic to metaphysical claims that the knowledge argument involves is complex and controversial, but it is not a non-starter. Yet it would be a non-starter if we construed Mary's progress in terms of her gaining mere knowledge, instead of knowledge$_M$ (Alter

---

the expert's knowledge on some topic is complete and not subject to correction. However, science is always subject to progress and changes. For this reason, it is better to say that one uses the concept with mastery, rather than non-deferentially.

2003: 488).

Alter argues that, even in non-phenomenal cases, deferential possession of concepts is not sufficient for a priori deducibility: someone who has the concept ***Prime Numbers*** under an impoverished conception of ***Prime Numbers*** would not be able to deduce that there are infinitely many prime numbers, even if she has an ideal reasoning capacity. So knowledge$_M$ seems to be more appropriate when aiming at to a priori deducibility, and for this reason, more appropriate to explain Mary's epistemic gain. After leaving the black and white room, Mary becomes an expert on $\boldsymbol{Red}_P$ just like her peers who lived outside the room. It is an open question what truths someone would be able to deduce from complete physical knowledge, if the knowledge in question would be knowledge under impoverished conceptions. It is not to be expected that deferential possession of concepts would allow Mary to a priori deduce phenomenal truths from physical truths. Would someone be able to deduce that water is H_2O if one has an impoverished conception of $\boldsymbol{H_2O}$? Or that heat is molecular motion if one has an impoverished conception of molecular motion? Those observations suggest that a priori deducibility abilities depend on the full-mastery of relevant concepts. The conclusion is that the phenomenal concept strategy must be reformulated in terms of non-deferential mastery of concepts.

Michael Tye (2009) responds to the concept-mastery objection. He argues that it has an unacceptable consequence for phenomenal concept theorists, who hold that Mary makes no genuine discovery when she leaves the room. According to Tye, the concept-mastery thesis entails that there is no epistemic gap. This is so because the truths that Mary learns inside the room are the same truths that she learns after she leaves the room. However, this criticism is too simple. It could be raised against every kind of type-B materialism, which holds that Mary learns new modes of presenting old facts. Not learning new facts, does not mean that Mary makes no epistemic progress. On the concept-mastery objection, acquiring mastery of concepts counts as epistemic progress. Tye's objection is no more than a flat denial of what Alter tries to make plausible.

The concept-mastery objection's proposal is that experience is a condition for concept-mastery but not for concept possession. This change in the commitments of the phenomenal concept strategy allows Mary to possess phenomenal concepts while still inside the room; it does not entail that she does not learn anything or that phenomenal concepts do not exist. It entails only a different explanation of her epistemic progress: released Mary gains mastery of the concept she already 'possessed without mastery' inside the room.

Even if concepts of experience satisfy Burge's criteria for deferential use, the argument from social externalism fails when facing Alter's alternative proposal for the experience-dependence of phenomenal concepts. I therefore conclude that the physicalist may respond to Ball and Tye's objection successfully.

3 Chalmers' master argument

At last we arrive at David Chalmers master argument against the phenomenal concept strategy. Chalmers (2010) advances his powerful argument in the form of a dilemma. His argument attacks the general commitments of the strategy irrespective of the specific account of phenomenal concepts to which different proponents of the strategy subscribe. Chalmers describes first our epistemic situation concerning consciousness. This is the totality of our beliefs and epistemic intuitions about consciousness: the fact that we think we are conscious, that we introspect the qualitative properties of our mental states, that we think that Mary gains new knowledge, that we think that zombies are conceivable, that we think that there is an epistemic gap between physical and phenomenal truths. Let 'E' be a sentence that describes our epistemic situation regarding consciousness. Let us further assume that a complex thesis C involving some notion of phenomenal concepts applies to human beings, thereby forming the *explanans* in a explanation of our epistemic situation regarding consciousness: C explains E. The explanation in this case is, of course, not causal. Rather, it is a metaphysical, reductive explanation. As Chalmers puts it: the 'explanation that makes transparent why some high-level truths obtain, given that certain low-level truths

obtain' (Chalmers 2010: 313). This means, in Chalmers' framework, that E must a be a priori deducible from C, which means, in turn, that E&¬C is not conceivable. Thus the inconceivability of E&¬C is a condition that any adequate explanation of our epistemic situation regarding consciousness must satisfy. Another relevant condition must be satisfied if the explanation of E in terms of C should also vindicate physicalism: C itself must be explicable in physical terms. Again, this means, in Chalmers' framework, that P&¬C is inconceivable, where 'P' represents the complete physical description of the world. The main contention of Chalmers' master argument says that these two conditions cannot be simultaneously satisfied: either the physicalist explains our epistemic situation regarding consciousness or she maintains her physicalist view.

Chalmers' argument explores the case of the zombie. This is a creature conceived by us as partially described by P&¬Q. It should be noted that Chalmers is not begging the question against the phenomenal concept strategist. The latter agrees with Chalmers that there is a gap at the epistemic level: P&¬Q is conceivable. But it does not follow automatically from this that P&¬Q is metaphysically possible. Moreover, Chalmers takes it as very plausible that the zombie is in an epistemic situation regarding consciousness very different from the situation in which we humans are. That is, the sentence describing the zombie is 'P&¬Q&¬E'. What about C? Does the zombie satisfy the description C? It seems not. For C metaphysically explains E; and if the zombie satisfies ¬E, she cannot satisfy C. After all, C&¬E is inconceivable or C does not explain E. The result is that the zombie is partially but inevitably described by P&¬C. (The total description is, of course, P&¬Q&¬C&¬E). But if P&¬C is conceivable, P does not explain C. The problem is that this excludes a physical explanation of C. On the other hand, if we want to preserve a physical explanation of C, then we accept that P&¬C is not conceivable. However, grating that P&¬C is not conceivable allows that zombies satisfy C. Nevertheless, the thesis C, besides being physically explicable, is supposed to explain our epistemic situation regarding consciousness. If zombies satisfy C and C explains our epistemic situation, then the immediate conclusion is that zombies share our epistemic situation. But as we have seen, zombies do not share our E, since E entails

having phenomenal states. The only conclusion is that, if P&¬C is not conceivable, thesis C cannot explain our epistemic situation, hence C&¬E is conceivable. So either C is physically explained or C explains our epistemic situation regarding consciousness.

Thus we have a dilemma:

(H1) If P&¬C is conceivable, then C is not physically explicable.

(H2) If P&¬C is not conceivable, then C does not explain our epistemic situation.

∴ Either C is not physically explicable or C does explain our epistemic situation.

The master argument aims at the conclusion that no version of the phenomenal concept strategy can maintain that phenomenal concepts explains our epistemic situation *and* are physically explicable. The strategy is in trouble because, according to Chalmers, any physicalist strategy that appeals to phenomenal concepts *needs* to endorse both explanatory goals. We can question whether the phenomenal concept strategy should really commit to Chalmers' characterization of its explanatory goals, that is, with the explanation of C in physical terms and with the explanation of Chalmers notion of epistemic situation E. This will be explored in this section.

The way to respond to this dilemma is to attack its fragile points. There are four possible ways (Chalmers 2010: 320) to attack the dilemma proposed: (i) accept that C cannot be physically explained but show that it does not affect the phenomenal concept strategy; (ii) accept that C cannot explain our epistemic situation but the phenomenal concept strategy still has force for it can explain a different notion of epistemic situation; (iii) P&¬C is false so zombies share our epistemic situation; (iv) deny the connection between a priori and reductive explanation.

I will leave aside options (iii) and (iv) and I will consider one point against each horn of the dilemma concerning the Chalmers' characterization of the explanatory goal of C (options (i) and (ii)). Proponents of the phenomenal concept strategy may say that Chalmers has inflated the explanatory project of the phenomenal concept strategy (Diaz-Leon 2010, Balog 2012a, 2012b).

Regarding the first horn of the dilemma, accepting Chalmers' preferred sense of explanation yields the result that the account of phenomenal concepts cannot be explained in physical terms, since P&¬C is conceivable, then P does not explain C. The physicalist may question whether the phenomenal concept strategy must really accept the explanatory goals described by Chalmers. Must phenomenal concept theorists provide an account C that is physically explicable? I think that they do not have to grant that much. The phenomenal concept theorist can respond that Chalmers has failed to produce a correct picture of the explanatory goals of the strategy. The phenomenal concept strategy aims at blocking anti-physicalist arguments. Nevertheless, in order to block the arguments, the physicalist does not need to provide a *complete* physicalist explanation of phenomenal concepts. Proponents of the strategy must only deliver an account of phenomenal concepts that is *compatible* with physicalism. The reaction points to Chalmers' failure in characterizing the appropriate commitments of the phenomenal concept strategy: the strategy must not close the epistemic gap between the phenomenal and the physical by simply offering a physicalist explanation of facts about consciousness. It suffices that the strategy presents a feature of phenomenal concepts, which is neutral regarding its ontological commitments. The task of the phenomenal concept strategy is, therefore, to explain *why* there is a gap and *how* phenomenal concepts are cognitively isolated from physical concepts. Those facts about phenomenal concepts must be *consistent* with physicalism, but they do not need to be physically explained. And that much, the phenomenal concept strategy delivers.

In Balog's account of phenomenal concepts this becomes explicit. The acquaintance relation is supposed to afford special epistemic access to the referent of the phenomenal concepts we deploy in introspection: acquaintance reveals the nature of the referent in a way that no other epistemic relation does. This special feature explains Mary's epistemic progress upon leaving the black and white room and it explains the difference between me and my zombie counterpart: post-release Mary and I have epistemic access via acquaintance to the 'real nature' of our phenomenal states and pre-release Mary does not. However, postulating acquaintance also seems to open a new explanatory gap, if we do not give a physicalist explanation

for acquaintance we are left with treating it as a primitive. Hence, acquaintance is left physically unexplained. Defenders of this strategy require it to be merely compatible with physicalism, not physically explicable (Balog 2012). No version of the phenomenal concept strategy must provide a physicalist explanation of phenomenal concepts. What they provide is an account that is compatible with physicalism.

Regarding the second premise of the argument a complete physical description of the world P cannot explain our epistemic situation regarding consciousness because the conceivability of P&¬C would entail that zombies share our epistemic situation: if a physical description P entails a priori the thesis C, then a creature which is physically identical to me (my zombie) would also satisfy C. Having C should be what explain our epistemic situation. If zombies also have C, then they share our epistemic situation. However, according to Chalmers, sharing our epistemic situation with zombies is highly implausible. Hence, C cannot explain our epistemic situation (C&¬E).

There is an important point in the second horn of the dilemma (H2) that is Chalmers' special notion of 'epistemic situation':

> Two individuals share their epistemic situation when they have corresponding beliefs, all of which have corresponding truth values and epistemic status (Chalmers 2010: 316)

Regarding the second horn of the dilemma, Diaz-Leon (2010) charges Chalmers' with proposing again an incorrect characterization of the explanatory goal of the phenomenal concept strategy regarding our epistemic situation concerning consciousness. Indeed, if a complete physical description (P) entails an account of phenomenal concepts (C), then it is possible that a creature physically identical to me also satisfies an account of phenomenal concepts. If zombies have phenomenal concepts, then they either share our epistemic situation regarding consciousness or C cannot explain E. Now it all comes down to the specific notion of epistemic situation employed by Chalmers. Let us consider his definition again:

(E_1) The epistemic situation of an individual includes the truth values of their beliefs and the epistemic status of their beliefs and the inferential disconnection between phenomenal and physical beliefs.

Now, the conceivability of zombies (P&¬Q) is what creates a problem for the second horn of the argument. (E_1) prevents us from sharing the epistemic situation with zombies, because it requires as part of the epistemic situation having true beliefs about our phenomenal states. In order for one to have true beliefs about one's phenomenal states, one must have phenomenal states. Because I have phenomenal states, but my zombie does not, zombies do not share our epistemic situation even if they satisfy C. So C cannot explain our epistemic situation as formulated in (E_1).

Nevertheless, the proponent of the C may claim that C does not have to explain our *entire* epistemic situation regarding consciousness. So what would be the adequate characterization of the strategy's explanatory goal? The explanatory task of the phenomenal concept strategy is clear from the beginning, it must *only* explain the inferential disconnection between phenomenal truths and physical truths, it does not imply that one has phenomenal states to justify one's phenomenal beliefs (as required by Chalmers). So we can characterize the epistemic situation not as (E_1) but as (E_2):

(E_2) There is an inferential disconnection between P and Q.

The second horn of the dilemma claims that zombies satisfy C, since physical truths entail C. C entails a priori that there is an inferential disconnection between phenomenal and physical concepts (C→E_2). Hence, a creature physically identical to us with no phenomenal states would also instantiate such a gap, the creature would also be unable to infer phenomenal truths form physical truths. This is why it is possible that zombies share our epistemic situation, because C does not entail our entire epistemic situation (E_1), rather C entails only the epistemic gap (E_2).

According to Diaz-Leon (2010), this point becomes more clear when we focus on one specific version of the strategy. Noteworthy is the recognitional account proposed by Hill and McLaughlin (1999). According to them, what

explains the conceptual isolation of phenomenal concepts is the fact that both concepts play very different psychological roles. If we characterize these roles in functional terms, then zombies would also share those roles with us. Consider that, while I have phenomenal concepts, my zombie-twin has only quasi-phenomenal concepts. Quasi-phenomenal concepts are functionally like my phenomenal concepts (but not phenomenally like mine). There is also an epistemic gap between the zombie's quasi-phenomenal concepts and physical concepts, since they involve different psychological faculties, they will not be a priori connected. The zombie, like us, cannot infer a priori quasi-phenomenal beliefs from physical beliefs. This shows that we share our epistemic situation with zombies. Epistemic situation understood as the inferential disconnection between physical the phenomenal beliefs.

> Therefore, we can conclude that, if we understand the epistemic gap as an inferential disconnection between physical and phenomenal beliefs, then there is no evidence that C might hold without the epistemic gap holding. In particular, Chalmers' alleged counterexample, namely, zombies, is not a real case of beings that satisfy C but not the epistemic gap, since there is no relevant epistemic gap that they fail to satisfy, even if they do not instantiate all the aspects of our epistemic situation. Hence, zombies do not represent a problem for the claim that C can explain the epistemic gap. (Diaz-Leon 2010: 945)

The existence of zombies does not represent a problem for the phenomenal concept theorists if they consider the proper interpretation of our epistemic situation as (E_2) instead of (E_1). The explanatory goal of the strategy is to allow that C explains only the epistemic gap (the inferential disconnection) and not our entire epistemic situation (inferential disconnection plus the presence of phenomenal states).

In fact, Diaz-Leon (2010) points to a problem raised by the adoption of Chalmers' preferred sense of epistemic situation (E_1). Chalmers proposes that the definition of epistemic situation (E_1) entails that the subject in question has phenomenal states that justify their truth beliefs. For this reason we cannot share (E_1) with zombies, for zombies do not have phenomenal states. So, by definition (E_1) entails Q a priori (where Q is a phenomenal

state). Diaz-Leon formulates an undesirable consequence of (E_1) (2010: 947):

1. P a priori entails C
2. C a priori entails E_1
3. E_1 a priori entails Q

$\therefore$ P a priori entails Q

If we accept Chalmers' picture of the explanatory role of the phenomenal concept strategy and the second horn of his argument, the result is this: for physicalism to be true, a thesis C must be reductively explained by a complete physical description (P→C). Mutatis Mutandis, our epistemic situation (E_1) can only be reductively explained by an account of phenomenal concepts (C) if C entails E a priori. The definition of epistemic situation (E_1) includes possession of phenomenal states (Q), so if one is in an epistemic situation (E_1), one has phenomenal states (Q). If all this is true, then a complete physical description P must entail Q. Then there is no epistemic gap left to be explained. The conclusion is that (E_2) is far superior than (E_1) as a suitable candidate to characterize our epistemic situation. Besides, (E_2) does not run into the same problems as (E_1), since the former is perfectly compatible with us sharing our epistemic situation with zombies, while the latter is not. Moreover, characterizing the epistemic situation by requiring the presence of phenomenal states leads to the conclusion that there is no epistemic gap. This is not an acceptable explanatory goal: that of explaining a non-existent epistemic gap.

Hence, Chalmers' dilemma is not a conclusive argument against the phenomenal concept strategy. First, the phenomenal concept theorist is not obliged to grant that a thesis C must be physically explicable. The task of the phenomenal concept strategy is to block anti-physicalist arguments by simply offering an alternative explanation of the epistemic gap which is compatible with physicalism. Second, even if we grant the first horn of the dilemma, the phenomenal concept strategy still has a reason to resist the conclusion of the second horn. They should only claim that Chalmers' characterization of epistemic situation is again not something that the strategist

should commit to explain. Again, the phenomenal concept strategist is committed to explaining the epistemic gap of the anti-physicalist arguments and not our entire epistemic situation regarding consciousness.

Chapter 5

Conclusion

So far I have argued in defense of the phenomenal concept strategy, which provides a physicalist solution to problems posed by epistemic arguments against physicalism, viz. the conceivability argument and the knowledge argument. My first task in this work was to provide a more precise definition of physicalism in Chapter One. My goal was to arrive at a definition of physicalism that is threatened by the anti-physicalist arguments. The background question that structured the discussion in Chapter One was: which formulation of physicalism do the conceivability argument and the knowledge argument attack? The various attempts to respond consisted in formulating the minimal commitments a theory must meet to be a physicalist theory. To evaluate these attempts we have to distinguish three questions: The base question, the scope question and the relation question. With respect to the relation question we found that supervenience physicalism is the most adequate account for physicalism. Moreover, supervenience physicalism implies the entailment thesis, which entails a formulation of a priori physicalism, i.e. the thesis that physical truths metaphysically necessitate phenomenal truths a priori. Many anti-physicalists think that this is what a physicalist theory should accomplish. With that formulation in hand, we saw, in Chapter Two, how the principal anti-physicalist arguments confront the metaphysical doctrine of physicalism with apparent facts about consciousness. Both anti-physicalist arguments depart from premises about the inferential disconnectedness between physical and phenomenal truths. Re-

garding the conceivability argument, the disconnectedness is present in (P1) which questions the falsity of the physicalist conditional by granting the conceivability of zombies (P&¬Q). Regarding the knowledge argument the lack of the requisite connection between physical and phenomenal truths appears in Mary's inability to deduce the latter from the former. Both arguments depart from the so-called epistemic gap to the effect that P does not entail Q a priori, hence zombies are conceivable and Q is not deducible from P. Up to this point, the physicalist can agree with the anti-physicalist. Although some physicalists reject the first premise of the argument, I counseled against following their example. Physicalists grant the inferential disconnection between P and Q but reject the next step of the anti-physicalist arguments: the step that infers an ontological gap from the epistemic gap. This step is grounded in the so-called rationalist interpretation of the two-dimensional semantics.

Chapter Two offered a detailed exposition of the two-dimensional framework which allows the distinction of at least three dimensions of conceivability. In Chalmers' interpretation of the two-dimensional semantics, at least one kind of conceivability should entail metaphysical possibility, which is what the proponent of the conceivability argument requires to argue against physicalism. The initial notion of conceivability is negative conceivability: S is conceivable if and only if S cannot be a priori ruled out. The distinction between two dimensions of conceivability should allow the proponent of the conceivability argument to accommodate common counterexamples to the link between conceivability and possibility as the truth and falsity of the Goldbach Conjecture. Ideal conceivability abstracts away from our cognitive limitations, while prima facie conceivability does not. This accommodates the well-known counterexamples to the simple inference from conceivability to possibility: the Goldbach conjecture is prima facie both conceivable as false and as true, but can ideally be conceived only either as false or as true. This leads to the conclusion that prima facie conceivability is not a good guide to metaphysical possibility.

Another counterexample to the link between conceivability and possibility involves the so-called Kripkean modal hybrids. Since the sentence 'Water is H_2O' is a posteriori it is conceivable as false even though it is metaphys-

ically necessary. How is that possible? This calls for one more distinction, primary vs. secondary conceivability which reflects two kinds of possibilities (primary vs. secondary). In the two-dimensional semantic framework, primary possibility is tantamount to considering a world as actual, whereas secondary possibility is tantamount to considering a world as counterfactual. This motivates a distinction between two kinds of intensions (also primary and secondary). Primary intensions are functions, whose inputs are worlds considered as actual (primary possibilities) whereas secondary intensions are functions, whose inputs are worlds considered as counterfactuals (secondary possibilities). The distinction between primary and secondary conceivability and possibility explains why modal hybrids *seem* to serve as counterexample to the link between conceivability and possibility. Considering a XYZ-world as actual, 'Water is XYZ' is primarily conceivable as true, hence it is primarily possible that 'Water is XYZ' is true. Nevertheless, there is no secondary possibility that 'Water is XYZ' is true since in the actual world 'water' designates H_2O. The conclusion is that primary conceivability is a good guide to primary possibility, but not a good guide to secondary possibility.

The kind of possibility relevant to the conceivability argument is secondary possibility, not primary possibility. This allowed us to refine the conceivability argument. There is no controversy about stepping from primary conceivability to primary possibility. However, how can we proceed from primary possibility to secondary possibility? In fact, primary conceivability is not a good guide to secondary possibility in *statements involving theoretical terms*. However, the two-dimensional semantics allows for the entailment between primary possibility to secondary possibility in statements involving phenomenal terms and microphysical terms. This has to do with the fact that the terms are semantically stable, or in other words, they have coinciding primary and secondary intension.

If some linguistic expression Q has coinciding primary and secondary intensions, then the same possibilities will be true in Q, since intensions are defined as functions from possibilities to truth-value. If Q has the same truth-value regardless of the possibility in which Q is evaluated, then there is no gap between primary and secondary possibility. In order for the sentence P&¬Q to be secondarily possible, both P and Q must have coinciding

primary and secondary intensions.

Among the semantically stable terms are phenomenal terms like 'pain' and microphysical terms like 'H_2O'. We have evidence that primary and secondary intensions differ when the intension of the linguistic expression and its extension seem to come apart. However, when the linguistic expression designates a phenomenal property, there can be no conceiving of phenomenal properties as different from their appearance, since it is the actual appearance of phenomenal properties that makes them what they actually are. There is a clear disanalogy between the appearance of contingency of phenomenal terms and the appearance of contingency of physical terms. For there is a strong dissociation between appearance and reality in the case of 'water' and 'heat', on the one hand, which does not occur in the case of conscious phenomena such as pain, on the other hand. This is how an advocate of the conceivability argument attempt to close the gap between primary possibility and secondary possibility in the case of phenomenal properties.

The knowledge argument has been exposed and critically assessed in Chapter Two. We have seen two formulations of the knowledge argument and possible responses to it. It became clear that the knowledge argument can only work if we consider a sort of a priori physicalism, that is, a version of physicalism according to which the psychophysical conditional $P \rightarrow Q$ is a priori. If the conditional is a priori, then Mary must be able to deduce Q from P. But she is apparently not. The sort of physicalist response we have explored is the one that argues for an a posteriori status of the conditional, and hence an a posteriori kind of physicalism. This can be offered as a solution for both arguments. If the conditional is a posteriori, then we block the link between conceivability and possibility and we explain Mary's ignorance in terms of lacking a concept. However, as we have seen in Chapter Two and Three, there is reason to resist to the a posteriority of the conditional or the identity.

The sort of physicalist response to both conceivability and knowledge arguments which provides an a posteriori account of physicalism mobilizes phenomenal concepts as the center pieces of an explanatory strategy. The phenomenal concept strategy recognizes the epistemic gap. However, the way to explain the epistemic gap is not by inferring an ontological gap, but

by explaining it as a conceptual gap. They relocate the gap by arguing that the gap is not ontological but conceptual. Physicalists who choose this kind of response argue that phenomenal concepts, in virtue of their experience-dependence character, are conceptually isolated, i.e. not a priori connected to physical concept. But, although not a priori connected, they co-refer to physical properties, hence they are a posteriori connected to physical concepts. In order for the phenomenal concept theorist to account for a satisfactory physicalist response to the conceivability argument and the knowledge argument, he must deliver an account that distinguishes specific facts about the nature of phenomenal concepts (Q) which assures their conceptual independence of and their possible a posteriority with physical concepts (P) while not running into Kripkean resistance to the identification of P and Q and the two-dimensional problems for the conditional between P and Q. All exponents of the strategy agree on the perspectival character of phenomenal concepts. They, however, disagree regarding the nature of phenomenal concepts. Each version assigns specific facts about concepts that should explain the perspectival character and account for the aposteriority of physicalism. Thus any account of phenomenal concepts must respond to two questions:

(Q1) Why are phenomenal concepts conceptually independent of physical concepts?

(Q2) How can the psychophysical identification or conditional be a posteriori if phenomenal terms are semantically stable?

I have explored how three versions of the phenomenal concept strategy face with respect to (Q1) and (Q2). Perry's indexical account of phenomenal concepts does not respond to (Q2) since it does not even address the question of a posteriori connectedness, it is without response to the conceivability argument. However, Perry does provide a response to the knowledge argument by arguing that Mary's epistemic situation is analogous to indexical epistemic situations in general. Phenomenal concepts have the kind of context-sensitivity that is characteristic of indexical concepts. This explains the lack of a priori deducibility of Q from P. Likewise, one cannot in general

deduce subjective truths (beliefs involving indexical contents) from objective truths (beliefs free of indexical). Vis-à-vis the knowledge argument, Perry's two dimensions of contents seem to do the work. However, we seek an account that argues for the a posteriority of physicalism. So Perry's theory cannot serve the purpose.

Loar's recognitional account of phenomenal concepts claims that phenomenal concepts are like recognitional concepts which are formed in virtue of the subject's abilities to discriminate and to re-identify the same kind of object. Recognitional concepts work like type-demonstrative concepts, in that they enable us to identify, discriminate, classify or perceive an object of *that* kind without the mediation of any description. Because of the recognitional character of these concepts, they refer without any descriptive mediation. This already marks the difference between phenomenal concepts and physical concepts. It also explains why the concepts are independent: They have *direct* reference-fixing mechanisms.

Loar's response to (Q2), the question about how to reconcile a posteriori physicalism with semantic stability, consists in offering an alternative explanation for the semantic stability of phenomenal concepts. In fact, he delivers yet an independent reason for the stability of phenomenal concept. Because phenomenal concepts are not cognitively tied to physical concepts, their identification is a posteriori. For Loar, it is sufficient to appeal to the psychological difference of phenomenal concepts and physical concepts to explain the a posteriority of physicalism without running into the difficulties generated by Kripke's argument.

The last version of the strategy considered was Papineau's constitutional account of phenomenal concepts. Papineau thinks that phenomenal concepts work like perceptual concepts, which, in turn, work like *stored sensory templates*. New templates are activated when we have a perceptual encounter and we accumulate information about the referents of our perceptual states. The application of phenomenal concepts is made through sensory templates, since we use stored sensory templates to think about our own experience. When we employ a phenomenal concept, a sensory template of the corresponding experience is activated by the experience itself. This allows for a direct reference-fixing mechanism. Phenomenal concepts

use experience to refer to the experience itself. The idea that experience constitutes the phenomenal concepts allows to respond to both questions (Q1) and (Q2): First, the cognitive isolation of phenomenal concepts is due to their use-mention character. Secondly, because the experience, which is the referent of the phenomenal concept, also fixes its reference, phenomenal concepts are cognitively isolated from physical concepts. They can thus be a posteriori connected to physical concepts even though they are semantically stable

After having considered these specific accounts of phenomenal concepts, we conclude that the phenomenal concept strategy has sufficient resources to argue for the a posteriority of physicalism. The special character of phenomenal concepts allows for P and Q to be a posteriori connected. Based on these consideration, I have tried to defend the strategy from well-known objections such as Stoljar's from the a priori synthesizable, Tye's and Ball's objection from social externalism and Chalmers' master argument.

Regarding Stoljar's objection, I have concluded that his distinction of a priori and the a priori synthesizable fails to block the strategist's treatment of the conceivability argument. There is a relevant disanalogy between the psychophysical conditional and the a priori conditionals mobilized by Stoljar. In every case that Stoljar presents, there is an adequate response from the phenomenal concept strategist. Moreover, Stoljar's objection to the treatment of the knowledge argument also fails. He considers a case in which experienced Mary possesses phenomenal concepts and nevertheless, cannot infer phenomenal truths from physical truths. My objection is that his case fails to capture Mary's situation. Experienced Mary possesses phenomenal concepts, but her knowledge about the application of phenomenal concepts belongs to the physical antecedent of the psychophysical conditional, not to the consequent. With an impoverished antecedent of the conditional, it is no surprise that Mary cannot make the relevant deductions.

Concerning now Ball's critical approach, there is also a disanalogy between statements involving the possession of phenomenal concepts and that of non-phenomenal concepts. Because of this disanalogy, Burgean arguments that may apply to ordinary concepts do not apply to phenomenal concepts. However, even if the phenomenal concept theorist wants to grant that social

externalism is true of phenomenal concepts, Ball's argument cannot succeed. This is because the experience thesis can be reformulated in order to accommodate Ball's objection. Experience is not a condition for concept possession, as required by the experience thesis, rather, it is a condition for concept mastery. Mary's epistemic gain can be plausibly described in terms of concept mastery rather than in terms of concept possession.

I have concluded that the problem of the master argument is to inflate the phenomenal concept strategy's explanatory goal. The strategy need not explain phenomenal concept physically. It suffices to explain why phenomenal and physical concepts are cognitively isolated and how that is *compatible* with a physicalist ontology. That much the phenomenal concept strategy delivers. Also seconded by Diaz-Leon that Chalmers' characterization of an *epistemic situation* is not a commitment that the phenomenal concept strategy should accept.

References

Alter, T. (2001). Know-How, ability, and the ability hypothesis. *Theoria* (67), 229–239.

Alter, T. (2013). Social externalism and the knowledge argument. *Mind 122*(486), 481–496.

Alter, T. and R. Howell (2009). *A Dialogue on Consciousness*. Oxford: Oxford University Press.

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

Ball, D. (2009). There are no phenomenal concepts. *Mind 118*(472), 935–962.

Balog, K. (1999). Conceivability, possibility and the mind-body problem. *Philosophical Review* (108), 497–528.

Balog, K. (2009). Phenomenal concepts. In B. McLaughlin, A. Beckermann, and S. Walter (Eds.), *The Oxford Handbook of Philosophy of Mind*, Chapter 17, pp. 292–312. Oxford University Press.

Balog, K. (2012a). Acquaintance and the mind-body problem. In S. Gozzano and C. Hill (Eds.), *New Perspectives on Type Identity: The Mental and The Physical*. Cambridge University Press.

Balog, K. (2012b). In defense of the phenomenal concept strategy. *Philosophy and Phenomenological Research 1*(84), 1–23.

Beckermann, A. (2008). *Analytische Einführung in die Philosophie des Geistes*. de Gruyter Studienbuch.

Block, N. (Ed.) (1980). *Readings in the Philosophy of Psychology*, Volume 1. Harvard University Press.

Braddon-Mitchell, D. and F. Jackson (1996). *Philosophy of Mind and Cognition* (Second ed.). Blackwell.

Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy 4*, 73–121.

Chalmers, D. (1996). *The Conscious Mind*. Oxford University Press.

Chalmers, D. (2002). Does conceivability entail possibility? In T. Gendler and J. Hawthorne (Eds.), *Conceivability and Possibility*, pp. 145–200. Oxford: Oxford University Press.

Chalmers, D. (2004). Epistemic two-dimensionalism. *Philosophical Studies 152*, 153–226.

Chalmers, D. (2006). The foundations of two-dimensional semantics: Foundations and applications. In M. Garcia-Carpintero and J. Macia (Eds.), *Two-Dimensional Semantics*. Oxford University Press.

Chalmers, D. (2010a). *The Character of Consciousness*. Oxford: Oxford University Press.

Chalmers, D. (2010b). Phenomenal concepts and the explanatory gap. In D. Chalmers (Ed.), *The Character of Consciousness*. Oxford University Press.

Chalmers, D. (2010c). The two dimensional argument against materialism. In D. Chalmers (Ed.), *The Character of Consciousnes*, pp. 141–205. Oxford University Press.

Crane, T. (2001). *Elements of Mind*. Oxford University Press.

Crane, T. and H. Mellor (1990). There is no question of physicalism. *Mind 99*(394), 185–206.

Dennett, D. (1991). *Consciousness Explained*. Little, Brown and Co.

Dennett, D. (2007). What RoboMary knows. In T. Alter and S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Diaz-Leon, E. (2008). Defending the phenomenal concept strategy. *Australasian Journal of Philosophy 86*(4), 597–610.

Diaz-Leon, E. (2010). Can phenomenal concepts explain the epistemic gap? *Mind 119*(476), 933–951.

Diaz-Leon, E. (2011). Consciousness, phenomenal concepts, and acquaintance. *Teorema 30*(1), 157–167.

Garcia-Carpintero, M. and J. Macia (Eds.) (2006). *Two-Dimensional Semantics*. Oxford University Press.

Goff, P. (forthcoming). Real acquaintance and physicalism. In P. Coates and S. Coleman (Eds.), *Phenomenal Qualities: Sense, Perception and Consciousness*. Oxford University Press.

Gozzano, S. and C. Hill (Eds.) (2012). *New Perspectives on Type Identity: The Mental and The Physical*. Cambridge University Press.

Hempel, C. G. (1980). Comments on Goodman's "Ways of Worldmaking". *Synthese 45*, 193–199.

Hill, C. (1997). Imaginability, conceivability, possibility and the mind-body problem. *Philosophical Studies 87*, 61–85.

Hill, C. and B. McLaughlin (1999). There are fewer things in reality than are dreamt of in Chalmers' philosophy. *Philosophy and Phenomenological Research 59*, 445–454.

Horgan, T. E. (1984). Jackson on physical information and qualia. *Philosophical Quaterly 34*, 147–183.

Howell, R. J. (2013). *Consciousness and the Limits of Objectivity: The Case for Subjective Physicalism*. Oxford University Press.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quaterly* (32), 127–36.

Jackson, F. (1986). What Mary didn't know. *Journal of Philosophy* (83), 291–295.

Jackson, F. (1998a). *From Metaphysics to Ethics*. Oxford University Press.

Jackson, F. (1998b). Reference and description revisited. *Nous* (32), 201–218.

Jackson, F. (2004a). Postscript on qualia. In P. Ludlow, Y. Nagasawa, and D. Stoljar (Eds.), *There's something about Mary*. MIT Press.

Jackson, F. (2004b). Why we need a-intensions? *Philosophical Studies 118*, 257–277.

Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and the Mental Causation*. MIT Press: A Bradford Book.

Kripke, S. (1972). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kripke, S. (1979). A puzzle about belief. In A. Margalit (Ed.), *Meaning and Use*. D. Reidel.

Levin, J. (2007). What is a phenomenal concept? In T. Alter and S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, pp. 87–110. Oxford University Press.

Lewis, D. K. (1966). An argument for the identity theory. *Journal of Philosophy 63*(1), 17–25.

Lewis, D. K. (1980). Mad pain and martian pain. In N. Block (Ed.), *Readings in the Philosophy of Psychology*, Volume 1, pp. 216–222.

Lewis, D. K. (1994). Reduction of mind. In S. Guttenplan (Ed.), *Companion to the Philosophy of Mind*, pp. 412–431. Blackwell.

Lewis, D. K. (1999). New work for theory of universals. In *Papers in metaphysics and epistemology*. Cambridge University Press.

Lewis, D. K. (2004). What experience teaches. In P. Ludlow, Y. Nagasawa, and D. Stoljar (Eds.), *There's Something about Mary*. MIT Press: A Bradford Book.

Loar, B. (1990). Phenomenal states. *Philosophical Perspectives 4*, 81–108.

Loar, B. (1997). Phenomenal states (second version). In N. Block, O. Flanagan, and G. Güzeldre (Eds.), *The Nature of Consciousness: Philosophical Debates*. MIT Press.

Loar, B. (2003). Qualia, properties, modality. *Philosophical Issues 13*.

Ludlow, P., Y. Nagasawa, and D. Stoljar (Eds.) (2004). *There's something about Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*. Cambridge: MIT Press.

McLaughlin, B. and K. Bennett (2014). Supervenience. In *The Stanford Encyclopedia of Philosophy*.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review* (83), 435–450.

Nemirow, L. (1980). Review of Nagel's Mortal Questions. *Philosophical Review 89*(473-7).

Nemirow, L. (1990). Physicalism and the cognitive role of acquaintance. In W. Lycan (Ed.), *Mind and Cognition: A Reader*. Oxford: Blackwells.

Nida-Rümelin, M. (1996). What Mary couldn't know. In T. Metzinger (Ed.), *Conscious Experience*, pp. 219–241. Exeter: Imprint Academic.

Nida-Rümelin, M. (2002). Qualia: The knowledge argument. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

Papineau, D. (1993). *Philosophical Naturalism*. Blackwell.

Papineau, D. (2002). *Thinking About Conciousness*. Oxford: Oxford University Press.

Papineau, D. (2007). Phenomenal and perceptual concepts. In T. Alter and S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Perry, J. (1979). The problem of the essential indexical. *Nous 13*(3-21).

Perry, J. (2001). *Knowledge, Possibility and Consciousness*. Cambridge, MA: MIT Press.

Putnam, H. (1973). Meaning and reference. *Journal of Philosophy 70*, 699–711.

Putnam, H. (1975). The meaning of 'meaning'. In K. Gunderson (Ed.), *Language, Mind and Knowledge*, pp. 131–93. University of Minnesota Press.

Russel, B. (1905). On denoting. *Mind 14*(56).

Ryle, G. (1949). *The Concept of Mind*. Chicago: The University of Chicago Press.

Schroeter, L. (2012). Two-dimensional semantics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Syntax and Semantics: Pragmatics*, pp. 315–332. New York: Academic Press.

Stalnaker, R. (1999). *Content and Context*. Oxford University Press.

Stalnaker, R. (2008). *Our Knowledge of the Internal World*. Oxford University Press.

Stoljar, D. (2001a). The conceivability argument and two conceptions of the physical. *Nous* (35), 393–413.

Stoljar, D. (2001b). Two conceptions of the physical. *Philosophy and Phenomenological Research* (62), 253–281.

Stoljar, D. (2005). Physicalism and phenomenal concepts. *Mind and Language 20*, 469–494.

Stoljar, D. (2009). Physicalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

Tye, M. (2000). *Consciousness, Color, and Content*. Ma: MIT Press.

Tye, M. (2009). *Consciousness Revisited: Materialism without Phenomenal Concepts*. MIT Press.

Tye, M. (2015). Qualia. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

Veillet, B. (2012). In defense of the phenomenal concepts. *Philosophical Papers 41*(1), 97–127.

Deutsche Zusammenfassung

Es ist die zentrale metaphysische These des Physikalismus, daß die Verteilung *aller* Eigenschaften durch die Verteilung der Teilmenge der physikalischen Eigenschaften bestimmt ist. Hauptziel dieser Dissertation ist die Verteidigung dieser These gegen zwei Argumente, nämlich das Vorstellbarkeitsargument und das Wissensargument. Beide Argumente konfrontieren die physikalistische These mit augenscheinlichen Tatsachen über unser Bewusstseinsleben, nämlichen den besonderen qualitativen Charakter bewußten Erlebens. Beide Argumente nehmen als Ausgangspunkt epistemische Prämissen, um auf die metaphysische Folgerung zu schließen, dass der Physikalismus falsch sei. Dieser Schritt wird durch eine inferentielle Verknüpfung zwischen Vorstellbarkeit und metaphysischer Möglichkeit gerechtfertigt.

Im Laufe dieser Dissertation untersuche ich eine Reihe physikalistischer Reaktionen, welche die Verbindung zwischen epistemischer Vorstellbarkeit und metaphysischer Möglichkeit blockieren soll durch Rückgriff auf den besonderen Charackter sogenannter phänomenaler Begriffe. Die Kluft zwischen phänomenalen und physikalischen Wahrheiten wird so nicht als eine ontologische erklärt sondern als eine begriffliche. Die *phänomenale Begriffsstrategie* (wie sie in der Literatur genannt wird) schreibt den phänomenalen Begriffen besondere Eigenschaften zu. Diese besonderen Eigenschaften erklären ihre kognitive Unabhängigkeit von physikalischen Begriffen. Allgemeines Einverständnis unter Physikalisten ist, dass die besondere Eigenschaft phänomenaler Begriffe darin besteht, dass sie *erfahrungsabhängig* sind. Es herrscht aber Uneinigkeit im Bezug auf die Frage, wie es dazu kommt, daß diese Begriffe erfahrungsabhängig sind. Für Vertreter der phänomenalen Begriffsstrategie muß jede Antwort, die hier gegeben wird, derart sein, daß die Verbindung zwischen physikalischen und phänomenalen Wahrheiten bestenfalls a posteriori sein kann. Dadurch sollen die Argumente gegen den Physikalismus unwirksam werden.

In der Dissertation untersuche ich drei Varianten der phänomenalen Begriffsstrategie, die mir besonders aussichtsreich zu sein scheinen. Ich zeige, daß alle drei Strategien über genügend Mittel verfügen, antiphysikalistischen Argumenten zu begegnen und Angriffe abzuwehren. In dieser Hinsicht ist die Strategie also stark. Ich zeige aber auch einige Schwächen auf hinsichtlich ihrer Versuche, den Charakter phänomenaler Begriffe aufzuklären.