Self-Involving Representationalism (SIR): A naturalistic Theory of Phenomenal Consciousness

Miguel Ángel Sebastián González

:

MIGUEL ÁNGEL SEBASTIÁN

SELF-INVOLVING REPRESENTATIONALISM (SIR): A NATURALISTIC THEORY OF PHENOMENAL CONSCIOUSNESS

Miguel Ángel Sebastián: *Self-Involving Representationalism (SIR): A naturalistic Theory of Phenomenal Consciousness*

PROGRAMA:
Cognitive Science and Language

SUPERVISOR:
David Pineda

TUTOR:
Josep Màcia

DEPARTAMENTO:
Departament de Lògica, Història i Filosofia de la Ciència

FACULTAD:
Facultat de Filosofia

UNIVERSIDAD:
Universitat de Barcelona

A mis padres y a mi hermano.

A los compañeros de Logos.

ABSTRACT

A naturalistic account of phenomenal consciousness is presented: Self-Involving Representationalism.

The first step for the project of naturalizing phenomenal consciousness is to make the project itself feasible. The purpose of the first part of this work is to provide a suitable answer to some arguments presented against this enterprise.

In the second part I will develop the pillars of the theory. According to Self-Involving Representationalism, phenomenally conscious mental states are states that represent a specific kind of *de se* content. This content can be naturalized in first-order terms by appealing to a certain sense of self: the sense of a bounded, living organism adapting to the environment to maintain life and the processes underlying the monitoring of the activity within these bounds.

RESUMEN

Se presenta una teoría naturalista de la consciencia fenoménica: Representacionalismo Ego-Involucrado (Self-Involving Representationalism –SIR).

El primera paso hacia una tería naturalista de la consciencia fenoménica es hacer el proyecto viable. El proposito de la primera parte es dar una respuesta adecuada a ciertos argumentos presentados en contra del proyecto.

En la segunda parte desarrollo los pilares de la teoría. De acuerdo con el Representacionalismo Ego-Involucrado, los estados fenoménicamente conscientes son aquellos que representan un tipo específico de contenido *de se*. Este contenido puede ser naturalizado apelando a cierto sentido de ego: el sentido de un organismo finito que se adapta al ambiente para mantener su vida y los procesos que subyacen a la monitorización y control de la actividad dentro de tales lindes.

ACKNOWLEDGEMENTS

I am *conscious* that a lot of people have helped to make this thesis possible. *There is something it is like for me* to think of each one of them, and it is something really wonderful. I would like to thank all of them for their contributions:

I consider myself privileged to have worked within the excellent academic environment provided by *Logos Research group in Logic, Language and Cognition*. All researchers, lecturers, and Ph.D. students there have positively contributed in many different ways to this work. For the invaluable, fruitful and friendly research environment they created I am thankful to Cristina Balaguer, Lorenzo Baravalle, Adrian Briciu, Fernando Broncano-Berrocal, Oscar Cabaco, Gemma Celestino, Marta Campdelacreu, Fabrice Correia, José Diez, Manel Durán, Ambrós Domingo, Manuel García Carpintero, Jordi Fernandez, Sanna Hirvonen, Carl Höefer, Max Kölbel, Mireia Lopez, Teresa Marques, Genoveva Martí, José Martinez, Giovanni Merlo, Andrei Moldovan, Paco Murcia, Joan Pagès, Francesc Pereña, Manuel Perez Otero, Sergi Oms, Laura Ortega, Chiara Panizza, Peter Pagin, Laura Pérez, Mirja Perez de Calleja, Andreas Pietz, Josep-Lluis Prades, Daniel Quesada, David Rey, Luis Robledo, Sonia Roca, Pablo Rychter, Sven Rosenkranz, Fiora Salis, Gonçalo Santos, Moritz Shulz, Stephan Torre, Guiliano Torrengo, Jordi Valor, Víctor Martín Verdejo and Dan Zeman. I am also grateful to Esa Diaz-Leon.

A special acknowledgement to my supervisor David Pineda, for helping me extract the ideas from my head and organize them; for motivating me with very nourishing discussion; and for his sharp comments, which have improved my work so much. I would like also to thank Josep Macià, my tutor, for his support. I wish to acknowledge those who have read and commented substantial portions of the dissertation: Marc Artiga, Marta Jorba, Ekain Garmendia (who is also responsible for the good fortune I had to end up in Logos). I am grateful to Dan Lopez de Sa for so many fruitful and entertaining discussions (the whole discussion on vagueness is enormously indebted to him) and to Manolo Martinez, not just for reading and commenting on my entire thesis but also for patiently responding to each one of my doubts –at least two a day, one with each coffee, for the last few years.

I have also profited from my discussions with audiences in Barcelona, Bergamo, Birmingham, Bochum, Geneva, New York, Taipei, Toronto, and St. Andrews.

As I worked towards my Ph.D. I had the opportunity to visit the universities of Warwick, National Yang Ming University (Taipei) and NYU (New York). My research benefited from discussion with a number of proffessors and students:

In Warwick: Naomi Eillan, Stephen Butterfill, Guy Longworth, Louise Richardson, Johannes Roessler, Pietro Salis, Matthew Soteriou and Keith Wilson.

In Taipei: Allen Y. Houng and his students.

In New York: Jake Berger, Richard Brown, Marilia Espiritu Santo, Uriah Kriegel, Farid Masrou, Myrto Mylopoulos, Hakwan Lau, David Pereplyotchik, Jesse Prinz, Jennifer Renee Corns, David Rosenthal, Dan

Shargel, Jonathan Simon, Ekathrina Vostrikova and, especially, Ned Block.

I am very grateful to Sandra Angles and Oscar Cabaco for helping me with the outline in Catalan and to Christine Schmiedel for English help.

I want to thank my friends for their support, affection and the many enjoyable moments I have had with them during the time I have been writting this thesis.

My research was funded: from 2007 to 2010 by the Commitee for the University and research of the department of Innovation, Universities and Company of the Catalunya government and the European Social Fund (Amb el suport del Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya); from 2007 to 2009 the Ministerio de Educación y Ciencia, research project HUM2006-09923 (Semantics and Pragmatics of Special Contexts); from 2010, the DGI, Spanish Government, research project FFI2009- 11347 (*Meaning, Translation and Context*) and the Consolider-Ingenio project CSD2009-00056.

I would like to thank my brother, Rubén Sebastián, for many things, among them, for his support with all the neurological issues presented in this work.

Finalmente quiero agradecerle a mis abuelas su cariño y dedicar la finalización de esta tesis a la memoria de mi abuelo, que desde pequeño no hacía más que animarme a estudiar. Seguí su consejo tan al pie de la letra que ya en sus últimos días más bien preguntaba cuando dejaría de hacerlo.

No creo que pueda agradecer con palabras a mis padres su apoyo, comprensión y confianza incondicionales, así como el enorme sacrificio que han hecho durante tantos años por nosotros y sobre todo por todo el amor que nos han dado; por lo tanto, no diré nada de ellos.

CONTENTS

LIST OF FIGURES

Part I

INTRODUCTION

1 INTRODUCTION

Conscious experiences are probably the most familiar and at the same time puzzling aspects of our minds. We do not know anything more intimately than our conscious experiences while they are also one of the things that we understand most poorly. If one tries to isolate what is the subject matter by means of some sort of definition, one will realize how difficult it is; in fact, many have thought that any attempt to define consciousness in terms of more primitive notions is fruitless. Fortunately, it is reasonably easy to understand what we are talking about: it feels a certain way to undergo these experiences.

When we taste chocolate cake, or when we smell the aroma of recently brewed coffee, there is certain information being processed in our brains that leads us to go and buy a coffee, to continue eating or to think that this is too good to be wholesome. However, a description of this information processing does not, *prima facie*, completely characterize the situation; it is also accompanied by a 'subjective quality', it *feels* a certain way to smell the coffee or taste the chocolate.

Examples of conscious experiences are those one has while looking at the ocean or at a red apple; or drinking a glass of scotch or a tomato juice; smelling the coffee or the perfume of a lover; or listening to the radio or a symphonic concert. Further examples are bodily sensations such as pains, hunger pangs, orgasms, etc. Emotions also have a characteristic feeling, just consider the radiant feeling when you are happy or the languidness of depression. There is also a conscious experience associated with mental imagery; for instance, when one imagines a paradisiacal beach or remembers one's first kiss.

The case of conscious thoughts is more controversial, it is not clear to me whether there is any particular feeling associated with thought that goes beyond that associated with the mental imagery or the associated emotion that I have mentioned before; clearly thinking about a beach and thinking about a mountain *feels* different, but surely the mental imagery associated with these thoughts is different. It may well be that there is, nevertheless, something beyond such imagery when we consider different cognitive attitudes, for instance as Chalmers (1996, p. 10) notes, "desire seems to exert a phenomenological 'tug'."

This familiarity of conscious experiences and the difficulty to pin down the subject matter were nicely illustrated by Block's informal comparison between jazz and conscious experience when he appealed to Louis Armstrong's famous quote: "if you have to ask what jazz is, you'll never know." Something similar seems to be true of consciousness.

In the last thirty years or so, interest in consciousness from within philosophy of mind has stepped up enormously and with it the number of competing theories. Over this time, there has been a notable improvement in the measuring devices that allow us to study the mechanisms of our brains. These facts, combined with the still poor but increasing interaction between researchers from different fields within the cognitive sciences, have brought consciousness into the scientific debate. Many philosophers do not ignore the empirical evidence and some scientists take on board certain philosophical considerations.

In this framework, the present dissertation aims to shed some light on the study of consciousness; first by offering a critical review of some of the most relevant theories of consciousness from a philosophical standpoint, while also examining empirical evidence that lends itself to this discussion. I will furthermore suggest experiments that can empirically settle some debates. Finally, I will attempt to provide a theory of consciousness that intends to solve some of these problems.

In this introduction I will first present the problem that consciousness poses for materialism and a brief introduction to what is materialism.

The term 'consciousness' is used in our everyday language to pick out different phenomena. In section 1.2 I will try to clarify the target of this work: phenomenal consciousness.

In section 1.3 I make a conceptual distinction between two aspects of phenomenal consciousness: the qualitative character and the subjective character. This will be helpful for the project of developing a theory of consciousness: the qualitative character is what makes the experience the kind of experience it is and the subjective character is what makes the experience a phenomenally conscious experience at all. The relation between these two aspects and the problem of consciousness is also introduced.

In the last section, 1.4, I present the structure of this dissertation.

1.1 THE PROBLEM OF CONSCIOUSNESS

While I am writing this introduction I am having a rich conscious experience. I see the computer in front of me and a red apple close to it. I smell the aroma of my cup of coffee, feel the keys of my keyboard under my fingers and a soft pain in my knee that makes me think that I shouldn't have been playing handball for so long yesterday. I feel thirsty and decide to drink some coffee. While I approach the cup to my lips I smell its aroma and remember the delicious *ristreto* I used to take every morning in Italy last summer. I burn my tongue and decide to wait a minute to avoid burning my throat.

My experience has many properties: it happens at a given time, in a certain location, in virtue of it I avoid burning my throat, etc. Some of them are not very interesting, like the place or the time they happen. Others raise serious scientific issues, like how the information about the high temperature of the coffee is stored in my brain, how the information about different features within the same modality is integrated,[1] how the motor system is affected, etc. Despite the difficulty of the topic, there is nothing *prima facie* incomprehensible in these issues. Understanding how we have the ability to discriminate and integrate information, focus attention, report mental states, etc. constitutes what Chalmers (1996) calls the *Easy Problem*. They are *easy* problems because all we need to solve them is a characterization of the brain mechanisms that allow us to perform such a function. Cognitive sciences have made an enormous progress, specially in the last years, to understand the mechanisms underlying these processes, and no matter how complex or poorly understood they may be, these processes can be entirely

1 This is the binding problem. When we perceive a green square and a red circle, what neural mechanisms ensure that the sensing of green is coupled to that of a square shape and that of red is coupled to that of a circle? For a discussion of the relation between binding and consciousness see Revonsuo and Newman (1999).

consistent with our conception of the world as made out of matter and energy.

On the other hand, conscious experiences have a *subjective dimension*, undergoing them *feels* some way, or to borrow Nagel (2002)'s expression, *it is like something for the subject* to undergo them. I will call the *way it is like for the subject* to undergo the experience the phenomenal character of the experience. The phenomenal character gives rise to what Chalmers calls the *Hard Problem*. The Hard Problem is the problem of explaining how energy and matter give rise to consciousness: why do conscious experiences exist? How do they arise from physical systems? Why and how does physical processing in my brain gives rise to my rich inner life at all? A related problem is the question of the concrete character of conscious experiences: why is looking at a red apple like *this* and not like *that*?

Many philosophers and scientists have the impression that theories of consciousness have been unsatisfactory, in a rather principled and systematic way. That has led some philosophers to embrace a mysterianist postition and claim that understanding consciousness is beyond human capacities, that consciousness is cognitively closed to us (McGinn 1989): no matter how deep we reflect on the problem, how far our science goes, we cannot understand consciousness. I do find the problem of consciousness fascinating but I am not that pessimistic. The problem has motivated, nonetheless, different metaphysical views:

According to Cartesian dualism, minds, and consciousness with them, are not a part of the physical world; they are distinctively outside the natural order. The kind of dualism that Descartes's scholars had in mind is called "substance dualism": the mind and the body are different kind of substances. Nevertheless, this metaphysical view is not very popular nowadays. There is a more interesting form of dualism, called "property dualism", defended among others by Chalmers (1996). According to property dualism, there is just one kind of substance that has two distinct kinds of properties: physical properties and mental properties. In this case even if one concedes that mind and body are identical in our world one can resist materialism. A denial of property dualism entails the denial of substance dualism (the opposite is trivially false) because if all properties are physical, what makes an object non-physical?

I am a materialist. At a first pass, materialism is the metaphysical thesis that holds that everything in our world, and consciousness as part of it, depends on the physical things. Materialism is the topic of the next subsection.

1.1.1 *Materialism*

The thesis that everything in our world depends on physical things needs to be clarified. There are two questions that require being unpacked: what is the relation that holds between everything and physical things and what are physical things.

What is the relation that holds between everything and the physical?

The answer to the first question usually appeals to the notion of supervenience. Lewis (1986) nicely presented the idea of supervenience with the following example:

> A dot-matrix picture has global properties — it is symmetrical, it is cluttered, and whatnot — and yet all there is to the picture is dots and non-dots at each point of the matrix. The global properties are nothing but patterns in the dots. They supervene: no two pictures could differ in their global properties without differing, somewhere, in whether there is or there isn't a dot. (Lewis, 1986, p. 14)

The properties of the picture supervene on the properties of the dots; there cannot be differences in properties of the matrix picture without differences in the properties of the dots. Supervenience can be defined as follows:

> (Supervenience)
>
> A set of properties B supervenes on a set of properties A if and only if any two possible situations that are identical with respect to A-properties are also identical with respect to their B-properties.

For instance, if economical properties supervene on physical properties then any situation that is physically identical to the current situation is one in which there is a global crisis: any two possible situations that are physically identical (understood as indiscernible and not as numerically identical) are economically identical.

As presented above, the thesis of supervenience is underspecified. Depending of the kind of modality (logical, metaphysical or nomological) involved we can distinguish three different relations that can hold between A-properties and B-properties: logical supervenience, metaphysical supervenience and nomological supervenience. I have claimed that materialism is a metaphysical thesis; we are therefore interested in metaphysical supervenience:

> (Metaphysical supervenience)
>
> A set of properties B *metaphysically supervenes* upon another set A if and only if any two *metaphysically possible* situations that are identical with respect to A-properties are also identical with respect to their B-properties.

However, some philosophers have argued that there is an entailment between logical supervenience and metaphysical supervenience, at least in the case of consciousness.[2] The thesis of logical supervenience holds that:

> (Logical Supervenience)
>
> A set of properties B *logically supervenes* upon another set A if and only if no two *logically possible* situations that are identical with respect to A-properties are also identical with respect to their B-properties.

The constraints on what is logically possible are "largely conceptual" (Chalmers 1996, p. 35) and are tied to the notion of conceivability. To a first and very rough approximation, we can say that a situation is conceivable if we can think of it without logical contradiction. So, if we cannot conceive of two situations that differ with respect to

2 I will elaborate on this entailment in chapter 2.

B-properties without differing with respect to A-properties, then B-properties logically supervene on A-properties. An example suffices to illustrate the idea for the moment: a married bachelor is not logically possible, we cannot think of a bachelor that is married, it makes no sense. The idea of a bachelor who is married leads to a contradiction. Many philosophers maintain, and I will concede it, that there is no logical contradiction entailed by the idea of an individual that is physically indiscernible from me but lacks consciousness: a zombie. If this is right, then zombies are logically possible and consciousness would not logically supervene on physical properties.

The relation between metaphysical supervenience and logical supervenience is controversial. Some philosophers, like Chalmers (1996), maintain that conceivability, once properly refined as we will see in chapter 2, entails metaphysical possibility, at least in the case of consciousness. If this were true, then metaphysical necessity is just as strong as logical necessity (characterized by appealing to the refined notion of conceivability) and if consciousness does not logically supervene on physical properties then it does not metaphysically supervene on physical properties either.

Materialism, as I will understand it, is the thesis that maintains that all the properties of the actual world metaphysically supervene on physical properties. Therefore, materialists that accept that zombies are logically possible have to deny that they are metaphysically possible. Otherwise, there would be properties in our world that would not metaphysically supervene on the physical.

It will be useful for future purposes to present the materialist thesis in terms of possible worlds:

> Materialism is true in a possible world w if and only if any metaphysically possible world which is a *minimal physical duplicate* of w is a duplicate of w *simpliciter*.

The notion of 'minimal physical duplicate of w' is borrowed from Jackson (1994); by 'minimal physical duplicate of w' he means a world that is identical in all physical respects to w, but which contains nothing else.[3]

If materialism is true of the actual world, then economical properties metaphysically supervene on physical properties and any metaphysically possible world that is a minimal physical duplicate of the actual world would be a world where there is a global crisis. More broadly, if materialism is true of the actual world, then every possible world that is a minimal physical duplicate of it is a duplicate simpliciter of it.

Imagine that there were angels, non physical entities, in the actual world or that there were mental properties different from physical properties in the actual world. In such a case, a minimal physical duplicate of the actual world would not be a duplicate simpliciter, for it would lack angels and mental properties. So, if zombies were metaphysically possible, then there would be a metaphysically possible world that is a minimal physical duplicate of the actual world but

3 Alternatively, Chalmers appeals to the notion of positive duplicate:

> Materialism is true at a possible world w if and only if any world which is a physical duplicate of w is a positive duplicate of w.

Where 'positive duplicate' means a possible world that instantiates all the positive properties of the actual world, being a positive property "one that if instantiated in a world w, is also instantiated by the corresponding individual in all worlds that contain w as a proper part" (Chalmers, 1996, p. 40).

lacks consciousness and therefore materialism would turn out to be false. One of the arguments against materialism that I will present in chapter 2 maintains that if zombies are logically possible then they are metaphysically possible. I will try to resist that claim.

What are physical properties?

The second question that needs to be clarified is: what are the physical properties on which, according to materialism, every other property in our world metaphysically supervenes? Philosophers have commonly appealed to a theory of physics for that purpose (Feigl 1958; Smart 1978; Lewis 1994; Chalmers 1996): physical properties are the properties that a theory of physics tells us about.

This way of characterizing materialism is controversial; Hempel (1969) maintained that a theory-based formulation of materialism is either false or trivial. If, on the one hand, we appeal to our contemporary theory of physics, it is quite plausible that such a theory is wrong and consequently materialism would be false. If, on the other hand, we appeal to a future or ideal physics[4] we don't know what kind of properties such a theory would postulate and so the thesis of materialism remains obscure. How can we justify our belief in materialism as the thesis that any property in our world metaphysically supervenes on physical properties if we don't know which are the physical properties? We cannot predict that a future theory of physics will not include mental entities. In such a case, materialism would be trivially true: mental properties would be physical properties. According to such an hypothetical theory of physics, mental properties would be part of the set of properties on which any other property in our world metaphysically supervenes.

I will rely on a suggestion by David Lewis in between the two horns of the dilemma.[5] Lewis (1994) proposes to think of physical properties as those postulated by a future physics which is a *suitable improvement* over our current theory of physics. Lewis' idea is that we can assume that a future physics will be an adjustment and not a radical change of our current physics. The kind of properties that such a future physics will postulate will be 'relevantly similar'[6] to the ones that our current physics postulates. So, we can assume that the reasons we have now for believing in materialism, given what we nowadays consider to be physical properties, will keep being valid with the physical properties that the future physics tells us about.

However, this characterization cannot be a valid characterization of materialism. The reason is that some theories that we intuitively consider to be compatible with materialism, would fail to be so. For example, according to our current theory of physics, *being simultaneous simpliciter with event A* is not a physical property, it is not one of the properties that our theory of physics tells us about, nor a property that metaphysically supervenes on physical properties. But *being simultaneous simpliciter with an event A* is a physical property according to

4 For instance, Chalmers (1996, p. 33) maintains that physical properties are "the fundamental properties that are invoked by a complete theory of physics."

5 For different characterizations of materialism see Crook and Gillet (2001); Montero and Papineau (2005); Pineda (2006).

6 Explaining the conditions under which two properties are 'relevantly similar' in a non-question begging way is a complicate issue (Pineda (2006)). I will rest here on an intuitive grasp of the idea. Intuitively, the properties that the theory of relativity postulates are relevantly similar to the properties that Newtonian physics postulates, *redness* is not.

Newtonian physics and we do not want to say that Newtonian physics is not a materialist theory.

A correct characterization of the materialist position is a very difficult and interesting topic. However, for the purposes of this dissertation I think that I can rest on the Lewisian approach, which is very close to the one that some anti-materialists endorse. We may presume that a future theory of physics will postulate properties such as mass, space-time location, spin, electric charge, or maybe some 'relevantly similar' properties, but not properties such as having a financial crack, being a lion or dispensing beer; these properties supervene on the properties that the theory will postulate.

We will then assume that consciousness will not be part of a future theory of physics that might be regarded as a *suitable improvement* of our current physics. This assumption can be accepted by both materialists and dualists. The discussion between dualists and materialists can be characterized as whether or not consciousness metaphysically supervenes on the properties that this future theory of physics will postulate.

My aim in this work is not to make the case for materialism, but to search for an account of consciousness that is compatible with the truth of materialism: a naturalistic theory of consciousness. A theory is a naturalistic theory if and only if all the properties that the theory postulates are physical or metaphysically supervene on physical properties.

If one is interested in an approach to the problem that is compatible with materialism, as I am, one needs to provide an account of what makes it the case that having an experience is like something for its possessor in naturalistic terms.

There are interesting arguments for doubting that such an approach will succeed. I will go into the details of these arguments in chapter 2.

1.2 PHENOMENAL CONSCIOUSNESS

To a first approximation, *phenomenal consciousness* can be defined as the property of my experience that is responsible for the hard problem of consciousness, the property that seems to make consciousness not deducible from the physical facts.[7]

Conscious experiences have a subjective dimension, undergoing them feels some way; *it is like something for the subject* to undergo them. When I look at the red apple close to my computer, there is *something it is like for me* to have this experience. More precisely, there is a *redness way it is like for me* to have such an experience. I will call to the *way it is like for me* to undergo the experience the phenomenal character of the experience. It is not clear at all how something physical can give rise to the phenomenal character of my experience.

7 Kriegel (2009) rigidified the definition of phenomenal consciousness as: "The property F, such that, in the actual world, F causally produces (in the suitable reflective subject, say) the sense that the facts of consciousness cannot be deduced from physical facts"(ibid.,pp.3-4). This definition is, however, problematic. As I will argue in chapter 2 that 'the sense that the fact of consciousness cannot be deduced from phenomenal facts' is due to the concepts we deploy to refer to the phenomenal character of the experience and I think that what a theory of consciousness has to explain is precisely this character.

1.2.1 *Different Concepts of Consciousness*

In ordinary contexts, in everyday language, the term 'consciousness' is used to refer to different phenomena that give rise to different but maybe interrelated questions. None of these uses is specially privileged, but we need to get clear about the subject-matter of this work: phenomenal consciousness.

Creature Consciousness Vs. State Consciousness

Rosenthal (1986) made a distinction between creature consciousness and state consciousness. The former is a property of an organism or other relevant system (a suitable artificial system for example); the latter is a property of mental states of the being. Creature consciousness is the most common denotation of the term consciousness in folk language. However, we use the term 'conscious' to refer to different properties of the being.

To the very least, a conscious creature is a creature capable of sensing and responding to the environment. In this use 'conscious' is a synonym for 'sentience'. Different organisms respond to different elements of the environment; the amount of information they are sensitive to varies enormously. It is an open question where to draw the line between conscious and non-conscious creatures in this sense. For instance, plants respond to changes in the environment, but few people would ascribe them the property of being a conscious creature. On the other hand, mammals or birds are clearly conscious in this sense. There are nevertheless plenty of borderline cases, due to our nowadays partial knowledge of their sensory system and to the vagueness of the concept itself. Is an amoeba, or a shrimp, or a slug, conscious?

In a different sense, creature consciousness is often used as a synonym of wakefulness. Creatures able to sense exhibit also different degrees of alertness. In this sense, the predicate 'X is conscious' denotes a property of a being that is awaken and responsive as opposed to being in coma, under anesthesia or deeply slept.

Finally, there is a much more philosophically interesting notion of creature consciousness: the one directly associated with phenomenal consciousness. In this sense a creature is conscious if and only if there is something it is like to be this creature. The problem of phenomenal consciousness is presented as the problem of creature consciousness for instance by Nagel (2002), who introduced the famous phrase when he invited us to wonder 'what it is like to be a bat?' or more recently by Chalmers (1996) and his zombies. This is the sense of creature consciousness that is philosophically interesting: the sense under which a zombie is not a conscious creature, precisely because, *ex-hypothesi*, a zombie doesn't have phenomenally conscious experiences. We have the clear intuition that a virus lacks consciousness in this sense and that we are conscious in this very same sense, but what about an amoeba, a butterfly, a bull, a dog or a monkey? Which are the conditions that an organism or an artifact has to satisfy to instantiate creature consciousness in this sense?

The notion of (phenomenal) creature consciousness is parasitic on the notion of phenomenally conscious experience. A creature is phenomenally conscious if and only if it undergoes phenomenally conscious experiences. It is in virtue of having phenomenally conscious experiences that there is something it is like to be such a creature. It may well

be that it is indeterminate whether a creature is conscious in this phenomenological sense: would we say that a creature that had a unique phenomenally conscious experience is a conscious creature? How many are required?

Phenomenally conscious experiences are a kind of mental state:[8] a phenomenally conscious mental state. To a first approximation we can introduce phenomenally conscious mental states as follows:

> A mental state M of a subject S is phenomenally conscious if and only of there is something it is like for S to be in M.

That is, a mental state is phenomenally conscious if and only if it has phenomenal character. When I look at a red apple there is something it is like for me to see the apple; i.e. I am in a mental state that is phenomenally conscious.

There are other kinds of mental states like beliefs, doubts, desires, fears, etc. These classes of mental states are not exclusively disjunctive; for instance, a mental state can be a desire and be phenomenally conscious. There is something it is like for me to consciously desire that my mother is coming to visit me, something different from what it is like for me to consciously desire that Real Madrid wins the football league. I might have the Freudian desire to kill my father or fear of castration but there is nothing it is like for me to have this desire or fear. Some desires are phenomenally conscious and others are not.

Phenomenally conscious mental states have properties that distinguish them from other mental states. It is in virtue of these properties that being in this state is like something for its possessor. It is in virtue of these properties that the experience has the phenomenal character it has and a phenomenal character at all. I will call these properties, which phenomenally conscious experiences have and other kinds of mental states lack, phenomenal properties.A theory of consciousness has to provide a characterization of such properties in virtue of which a mental state is a phenomenally conscious mental state.

Rosenthal (Rosenthal, 1986, 1997) maintains that a conscious state is simply a mental state one is aware of being in. For instance, a conscious belief that FC Barcelona will win the league is to have such a belief and also to be simultaneously and directly aware that one has such a belief. Whereas Rosenthal's claim picks out a sense in which we say that a mental states is conscious, it is controversial that this is the sense we are interested in; namely, phenomenal consciousness. Block's distinction between phenomenal consciousness and access consciousness will be illuminating at this point to understand the controversy.

Phenomenal Vs. Access Consciousness

When I look at the red apple close to my computer I have a phenomenally conscious experience, I can report that the apple is red; I can take it and bite it or I can just believe that it is going to be a delicious dessert. My visual system generates a visual representation of the apple and the content of this representation is processed and made available to other systems like the one responsible for reports, actions or belief-forming; I

8 It is not my intention to analyze the notion of mental state here. In philosophy of mind the notion is generally taken to be basic and uncontroversial. I consider the basicness of the notion of mental state as a cautious start point for my dissertation, further work in this direction has to be done to secure this starting point.

am *aware* of this information. This can make us think that the function of consciousness is to somehow "enable some information represented in the brain to be used in reasoning, reporting and rational control" (Block 2002b, p. 160). Conscious mental states are mental states that satisfy this functional role. This is the sense in which 'consciousness' or 'awareness' are typically used in cognitive neuroscience.

In 'On a confusion about the Function of Consciousness' Ned Block (2002b) famously introduced the distinction between access and phenomenal consciousness. Block complains against current scientific practices in the study of consciousness for targeting the wrong phenomenon. Instead of addressing the problem of phenomenal consciousness they have targeted the relatively unproblematic cognitive problem presented above in the surrounding of phenomenal consciousness: access consciousness. Access consciousness is closely related but different to phenomenal consciousness. Being in a phenomenally conscious state feels some way; there is something it is like for the subject to be in that state. The hard problem of consciousness is due to phenomenal consciousness: how can it be that being in a state that satisfies such a role *feels this* way?

Access consciousness is first introduced by Block 2002b as follows:

> A state is A[ccess]-conscious if it is poised for direct control of thought and action. To add more detail, a representation is A-conscious if it is poised for free use in reasoning and for direct "rational" control of action and speech. (The "rational" is meant to rule out the kind of control that obtains in blindsight[9]). (ibid., p.168)

The detailed characterization of access consciousness may be complicated. Chalmers (Chalmers, 1996, 1997) noted that enumerating the kind of control required for access-consciousness can be avoided by defining it as "direct availability for global control" (Chalmers, 1996, p. 225).[10] The idea seems to be that a state is access conscious when its content is directly available "to bring to bear in the direction of a wide range of behavioral processes."

The concept of access consciousness is clearly different from the concept of phenomenal consciousness; the relevant question is whether or not these two concepts pick out two different properties. Block argues that they do. He claims that although normally both properties come together, a state can be phenomenally conscious without thereby being access conscious and the other way around.

The first example that Block provides in favor of access without phenomenal consciousness is the case of a functional duplicate, which is computationally identical to a person but lacking phenomenal consciousness. Block thinks that the case is conceptually possible, but this is enormously controversial. Many functionalists would deny this concep-

9 Blindsight is a condition of patients with damage in the first stage of their visual cortex. These patients present a scotoma or 'hole' in their visual field. They claim not to be able to see any stimuli when presented in this area of their visual field. However, they are able to guess with high accuracy about presented stimuli in this hole in a forced-response task. For a detailed presentation of the phenomenon see Weiskrantz (1986).

10 Block 2002c has pointed out that this interesting notion is also problematic and it has the disadvantage of being too general. It seems that access-consciousness would not be an information processing image of phenomenal consciousness, if an organism like a slug has phenomenal consciousness just in virtue of some mechanism of resources' control that the slugs commands.

tual possibility and furthermore Block's argument requires something stronger: metaphysical possibility.[11]

A better example is the case of what Block's call *superblindsighter*. A superblindsighter is an imaginary patient suffering blindsight who is trained to prompt himself at will in such a way that he guesses without being told to guess. The superblindsighter suddenly thinks 'there is an horizontal object to my right despite the fact that I cannot see it. I am going to grasp it'. Visual information of certain kind gives rise to a thought. She knows that there is an object in the area of her scotoma. She can even compare and report that there is something for him to see the object when it is in her visual field outside the 'blind' area and nothing when it is inside this area. If this were possible, then we would have an example of perceptual content that is access conscious but not phenomenally conscious.

The example is suggesting, especially after having seen a video of one of the last Weiskrantz's[12] patients. In this video, a patient with a scotoma that covers his whole visual field is able to walk through a corridor avoiding all kind of obstacles. However, one can doubt that the kind of control the blindsighter can have can go beyond the one this patient exhibits and there is a clear difference in the functional role of the state of the blindsight patient and those that we undergo; the information tracked by their visual states is available to many fewer processes and surely there is a functional difference between the way he processes information and the way we do.

The case of phenomenal consciousness without access consciousness is also controversial. One classical example is the absent driver (Armstrong 1981). Imagine you are driving on the way home and you are deeply concentrated, reflecting on your favorite philosophical puzzle or planning the meetings of the next day; you don't pay attention to the traffic lights. Suddenly you wonder whether it was red or green; you cannot even remember whether you made a left or a right at the intersection. The traffic light was directly in front of you; the information was processed and you safely drove through the intersection. According to Block, that is a case of phenomenal consciousness without access consciousness.

A similar example: I am writing my dissertation in the living room of my flat and suddenly the light goes off. I do not notice it because my laptop continues working on battery, but I realize that the soft noise made by the refrigerator has stopped. I hadn't noticed that noise before. The noise was not available for free reasoning before, so the content of the state was not access conscious.

As in the case of the absent driver there are two possible interpretations of this situation. The first one, the one defended by Block, is that my experience of the noise was phenomenally conscious despite the lack of access consciousness. The fact that I can remember what hearing the noise was like seems to be a good evidence of this option. On the other hand, Block's opponent would maintain that I didn't have a phenomenally conscious experience of the noise and therefore this situation is not an example of phenomenal consciousness without access consciousness. They can claim, for instance, that whereas my experi-

11 I will discuss Block's example in some detail in 2.1.2 when I properly introduce functionalism, the approach that I will favor.

12 Lawrence Weiskrantz was the discoverer of the phenomenon of blindsight. The video can be seen in: http://www.youtube.com/watch?v=nFJvXNGJsws.

ence of remembering the noise is phenomenally conscious, the previous experience was not. I find far more plausible Block's interpretation.

Block's argument is not completely compelling but at least it shows a clear conceptual distinction. However, if the two concepts are not merely two different concepts of the same property, but they pick out two different properties, how is the study of phenomenal consciousness possible? One of the most important tools in the study of consciousness are the reports of subjects, and we can report the content of a mental state only if this state is access conscious. Access and phenomenal consciousness usually come together but we should pay attention to the possibility of phenomenal consciousness without access. In the study of consciousness, reports of the subjects should be taken at face value unless we have good reasons for rejecting them. Reports are the first word but cannot be the last one if the cognitive access underlying reportability is not necessary for consciousness. I think we have good reasons for thinking that it is not, as I will argue in 5.2.2.[13]

Kriegel (2006) argues that access consciousness and phenomenal consciousness pick out two different properties: access consciousness is a dispositional property whereas phenomenal consciousness is a categorical property. Kriegel makes a very interesting proposal that vindicates the current study in cognitive neuroscience: phenomenal consciousness would be, according to him, the categorical basis of the dispositional property that access consciousness is.

Consider a dispositional property like fragility, a property of my wine glasses. Fragility is the property of that which can be easily damaged, broken or destroyed. My wine glasses manifest their fragility when they fall and often when I wash them. But they do not need to break to have the dispositional property; nothing has to actually happen for my wine glasses to qualify as fragile. The categorical basis of a disposition is the property in virtue of which it exhibits the manifest property of breaking when falling. In the example of the fragility of my wine glasses, the categorical basis is the molecular structure of the thin glass that constitutes them.

In the same sense, phenomenal consciousness is the categorical basis of access consciousness, according to Kriegel's proposal. It is in virtue of its phenomenal properties that the mental state is *accesible for free use*. The content of the mental state does not need to be actually accessed; there is no need for a report or a behavioral response for the mental state to qualify as access conscious, as there is no need for the glass to break to qualify as fragile. The relation between a disposition and its categorical basis further accounts for the relation between access and phenomenal consciousness. There can be cases of phenomenal consciousness without any content being accessed (the manifestation of the disposition that access consciousness is); in the same respect that my wine glasses have the molecular structure they have even if they don't fall and consequently break.

Consider the case of the soft noise my fridge is making. According to Kriegel, I am phenomenally conscious of the noise, but only when

13 Nothing in Block's argument prevents the success of a functionalist theory that does not rest exclusively on this kind of cognitive accesibility to processed information, for these are the terms in which access consciousness is defined. David Chalmers calls 'awareness' to the perfect functional correlate of phenomenal consciousness and offers an argument against the identification between awareness and phenomenal consciosness also based on modal dissociation; this argument, contrary to Block's one, is intended to be an argument against any materialist theory of consciousness. I will discuss the argument in detail in 2.1.3.

the noise stops do I access the content of this mental state. The categorical basis (phenomenal consciousness) of the dispositional property (access consciousness) was already there independently of whether the disposition is manifested or not, independently of whether I access the content or not.

In the case of access consciousness without phenomenal consciousness, the advocate of the categorical-dispositional approach for explaining the relation between phenomenal and access consciousness can embrace the reply above. She can deny that the property the superblindsighter instantiates is the same one as the property phenomenal consciousness is the categorical basis of. This line of argument would require a more detailed characterization of access consciousness that allows us to distinguish the kind of acces we have from the kind of access the superblindsighter has to the content of her visual perception. If superblindsighters are a nomological possibility, as the case of the Weiskrantz patient seems to suggest, then this is an empirical issue, but it is very plausible that the kind of mechanisms differ.

One interesting reason in favor of the categorical-dispositional approach is that it vindicates the current scientific research in consciousness. As many times in the history of science, scientists try to learn about a property by working around the dispositional property the former is the categorical basis of.[14] Scientists, by researching the causes of the disposition manifestation (reports and action control), learn about the disposition's categorical basis, the categorical basis of access consciousness: phenomenal consciousness.

Block has made an important conceptual distinction between phenomenal consciousness and access consciousness. Furthermore, the concept of access consciousness and the concept of phenomenal consciousness pick out two different properties, because access consciousness is a dispositional property whereas phenomenal consciousness is a categorical one, phenomenal consciousness is something occurrent. Kriegel has suggested that phenomenal consciousness is the categorical basis of access consciousness but I believe that the relation between access consciousness and phenomenal consciousness is a bit more complicated than what Kriegel suggests. I think that phenomenal consciousness is part of the categorical basis of access consciousness but not the whole story. There can be states that are phenomenally conscious but that are not *globally accessible*, and therefore lack the dispositional property that access consciousness is. I will present some empirical evidence in favor of this claim in 5.2.2 and further clarify the relation between access consciousness and phenomenal consciousness in 5.4.4.

An important question remains: what is the relevant categorical basis? To approach this question I need to introduce a further distinction between two different aspects of the phenomenal character.

14 Kriegel (2009) presents the example of hereditary properties. These properties are dispositional and have been investigated during centuries, but only very recently their categorical basis, DNA, has been discovered. He presents the example of Huntington's disease, a progressive neurodegenerative genetic disorder, which affects muscle coordination and leads to cognitive decline and dementia. Research into Huntington's disease has led to the discovery that the cause is a mutation of the cytosine-adenine-guanine (CAG) gene.

1.3 TWO ASPECTS OF PHENOMENAL CHARACTER: QUALITATIVE AND SUBJECTIVE CHARACTER.

At the very beginning of this chapter I introduced the notion of phenomenal consciousness by saying that having a phenomenally conscious experience *feels* some way or other; there is something it is like to be in a phenomenally conscious mental state.

Phenomenally conscious mental states have, whereas other kind of mental states lack, phenomenal properties. Coming back to the example of the experience I have while looking at the red apple, there is something it is like for me to have this experience and more precisely, a *redness way it is like for me* to have it. I called the *redness way it is like for me* to look at the apple the phenomenal character of the experience. In virtue of its phenomenal character, that an experience is the experience it is and a phenomenally conscious experience at all. A theory of consciousness has to provide a characterization of the properties that phenomenally conscious mental states have. The properties that make a mental state the concrete kind of phenomenally conscious mental state it is and a phenomenally conscious mental state at all.

This job can be divided into two tasks. To illustrate this idea let me present an example: imagine that we look for a theory of antennas. Such a theory has to explain what distinguishes different kinds of antennas; what distinguishes a dipole antenna from a Yagi or a parabolic one. But this theory also has to explain what the property that antennas have, and other objects lack, is; in this case, being a transducer of electromagnetic waves. Similarly, a theory of consciousness has to explain what properties phenomenally conscious states have that other states (non-phenomenally-conscious states) lack and what distinguishes different kinds of phenomenally conscious states between them.

The experience I have when I look at the apple and the one I have when I look at the golf course differ in character, but there is a property (nothing from what I have said prevents this property from being a highly disjunctive one) they both have and non-phenomenally-conscious states lack; a property that distinguishes phenomenally conscious mental states from non-phenomenally-conscious ones: there is something it is like to be in any of the former states.

So we can make a conceptual distinction between two different aspects of the phenomenal character (Levine 2001; Kriegel 2005, 2009); two aspects of the '*redness way it is like for me* to see the apple': the *redness* and the *for me-ness*. I will maintain that the first aspect accounts for the differences between phenomenally conscious mental states whereas the second one accounts for the differences between phenomenally conscious and non-phenomenally conscious mental states. I will call qualitative character the former and subjective character the latter.

A theory of consciousness has to characterize what it is like for the subject to undergo the experience; the phenomenal character. The qualitative character explains *what it is like* for the subject to undergo the experience, the concrete way it feels to undergo the experience. A theory of subjective character explains what it is like *for the subject* to undergo the experience. It abstracts from the particular way different experiences feel and concentrates on the problem of what makes it the case that a conscious experience feels at all, independently of the particular way it feels to undergo the experience.

The distinction between qualitative and subjective character is introduced as such by Joseph Levine:

> Let's take my current visual experience as I gaze upon my red diskette case, lying by my side on the computer table. I am having an experience with a complex qualitative character, one component of which is the color I perceive. Let's dub this aspect of my experience its "reddish" character. There are two important dimensions to my having this reddish experience. First, as mentioned above, there is something it's like for me to have this experience. Not only is it a matter of some state (my experience) having some features (being reddish) but, being an experience, its being reddish is "for me," a way it's like for me, in a way that being red is like nothing for–in fact is not in any way "for"-my diskette case. Let's call this the subjectivity of conscious experience.
>
> The second important dimension of experience that requires explanation is qualitative character itself. Subjectivity is the phenomenon of there being something it's like for me to see the red diskette case. Qualitative character concerns the "what" it's like for me: reddish or greenish, painful or pleasurable, and the like. From within the subjective point of view I am presented with these qualitative features of experience, or "qualia," as they're called in the literature. Reddishness, for instance, is a feature of my experience when I look at my red diskette case.(Levine, 2001, pp. 6-7)

Let me present these two aspects in a bit more detail.

1.3.1 *Qualitative Character*

The first aspect of the phenomenal character I wish to discuss is the qualitative character; it is the *way* it is like for me to have the experience. The qualitative character is what distinguishes among different kinds of experiences, what distinguishes my experience of looking at an apple from my experience of looking at the grass on the park or my experience of hearing the music in my mp3.

The qualitative aspect of an experience can be very complex. In the apple example I focused in one feature of my experience, its *redness*, but it is also *round-like* for instance. We can undergo more complex experiences, for instance, the experience of tasting a wine can be described as follows:

> Aromas of very high intensity, again the delicious red fruit, ripe and sweet black fruit also around the rear, moving the glass appears the timber but by the hand of the fruit well together, some licorice, toffee. Flavor entered silky and dense, well-balanced, fruity and flavorful taste with a long post that fills the mouth.[15]

All the qualities described constitute the qualitative character of the experience. If, while I taste the wine, I look outside my window in the dark and a Django Reinhardt's performance of *Shine* is playing in the

15 The description corresponds to the 2006 Abadia Retuerta Seleccion Especial (http://vt-castilla-y-leon.uvinum.com/abadia-retuerta-seleccion-especial-2006/abadia-retuerta-seleccion-especial_review-3125)

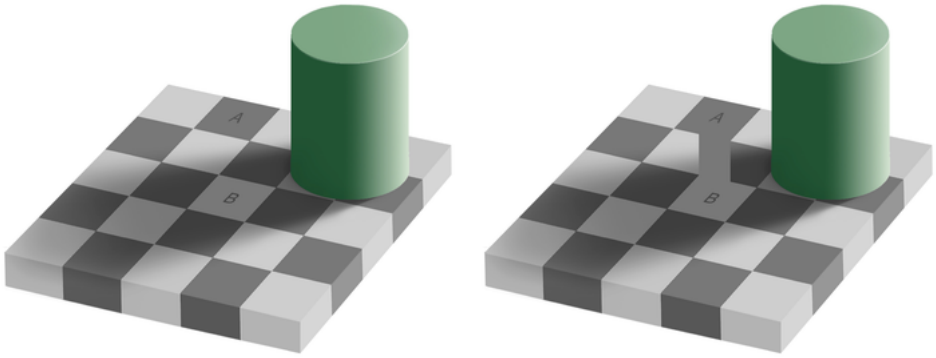

Figure 1: Adelson's Checker Shadow Illusion:
both squares A in B are the same color, although they seem to be different.

background, the overall qualitative character of my experience in this precise moment is the sum of the corresponding qualities: the sweetness, the denseness, the darkness, the rythmness, etc. But the qualitative character is not restricted to perceptual qualities. Every phenomenally conscious experience has qualitative features, for instance somatic or emotional experiences. The *stiffness* in my legs, the happiness or the stressfulness are also part of the qualitative character of my experience. Explaining the qualitative character is explaining in virtue of what a certain experience is the kind of experience it is.

My experience at any time has many different features. An interesting approach to the problem of explaining the qualitative character of the experience is to understand what distinguishes basic experiences. What distinguishes an experience as of certain shade of red from an experience as of a shade of green? So, from now on I will focus on basic qualities. A phenomenal quality is basic if and only if it is maximally specific, if it cannot be further decomposed into other phenomenal qualities. To a first approximation we can compare the experience you are having while looking at a red painting that occupies all your visual field with your experience looking outside the window, the former is more basic than the latter.

This approach may worry some readers. If phenomenal qualities interact with each other in a non mere additive way, then the qualitative character of the experience is not just given by the sum of instantiated phenomenal qualities but also by new qualities; which are the product of the combination of them. In this case the approach of focusing on basic qualities can be jeopardized.

For instance, it is well known that color perception in the foreground varies depending on the color in the background. Consider the Adelson's illusion presented in figure 1.[16] Check cells A and B are the same color, as the second image in the figure shows. However, they are perceived as being of a different color: what it is like for me to see A (greyness-A) is different from what it is like for me to see B (greyness-B). The reader might complain that this could be an indication that phenomenal qualities in experience cannot be explained independently of each other. We cannot explain greyness-B separately, because my experience having this quality might depend on the presence of other qualities.

16 http://en.wikipedia.org/wiki/Same_color_illusion

But many examples like this can be explained either at computational level or at physiological level without appealing to phenomenal qualities. Let's look to one such explanation. In the example of Adelson's illusion in the figure, the effect can be explained by the way the brain tries to determine the color of the checkerboard. Just measuring the light coming from a surface (the luminance) is not enough; for instance, a shadow can dim a surface. In that case the surface in shadow will reflect less light. In order to compensate for that effect, our visual system uses several tricks to determine the color of the surface. These tricks explain the effect of the illusion.

The first one is based on local contrast. A square of the checkerboard that is lighter than its neighboring checks is probably lighter than average, and the other way around. In the Adelson's illusion, the light square in shadow is surrounded by darker squares. Thus, even though the square is physically dark, it is light when compared to its neighbors. The dark square outside the shadow, conversely, are surrounded by lighter checks, so they look dark by comparison. Additionally, the shadows often have soft edges, while paint boundaries often have sharp edges. It is a good strategy to tend to ignore gradual changes and this way the visual system can determine the color of the surfaces without being misled by shadows.

Another interesting example is the interaction between moods and color perception. There is a cultural association, today and through the history, between depressed moods and dark colors. We have heard that when someone is depressed she is *black* or *grey*, and we often say to someone that is sad that she has to *brighten up*. When one is depressed everything looks greyer. Is this an example of interaction between phenomenal qualities? Do we have to appeal to the interaction between qualities to explain this effect? Bubl et al. (2010) showed that we do not. Using a technique called a pattern electroretinogram (PERG), they objectively measure the participant's ability to perceive contrast. They found a strong and significant correlation between the level of the depression and a decreased response in the PERG, suggesting that the more depressed the patient was, the less their retinas responded to the contrast pattern. If depression affects the retina response and the phenomenal character of visual experiences depends on the retina response then we do not need to postulate any interaction between phenomenal qualities in a strong sense that would prevent my analysis. So, if someone is depressed then the phenomenal character of her experiences changes because there is a change in the retina response. It seems reasonable to assume that such a change in the retina response would produce the same change in the phenomenal character of the experience even if the subject were not depressed.

Whether other examples of phenomenal interaction that the reader can think of can be explained in a similar fashion or not is an open question, but I take this starting point to be a compelling one.

I take the problem of explaining the qualitative character to be the problem of explaining what distinguishes basic experiences. What distinguishes an experience as of a concrete shade of red from an experience as of a different shade?

1.3.2 *Subjective Character*

The second aspect of phenomenal consciousness is the subjective character or *for me-ness*. If the qualitative aspects account for the differences in phenomenal character between experiences, the subjective aspect accounts for what all experiences have in common; in that sense it accounts for what makes an experience a conscious experience at all. The subjective character is what explains that being in a phenomenally conscious mental state *feels* at all. Phenomenally conscious mental states have, whereas non-phenomenally conscious states lack, *for me-ness.* This claim should be relatively uncontroversial. Being in a phenomenally conscious experience feels a certain way; for-meness is the property that binds phenomenally conscious experiences together and a theory of consciousness has to account for it. As a materialist it seems reasonable to maintain that there is a physical property that binds experiences together. As Tye (1996) claims:

> [T]here is something that unites all phenomenally conscious states: as noted earlier, each phenomenal state type is such that there is something it is like to undergo any possible token of that type. If there is no physical property that phenomenal states share, then the obvious conclusion to draw is that there is an aspect to the world that physicalism cannot capture.(Tye, 1996, p. 684)

I want to say a bit more about the *explanandum* in this introduction. One can be suspicious of the way I have presented the problem by appealing to for-meness and claim that the composition of the expression 'something it is like for the subject' is somehow artificial. One can complain that phenomenal character, there being something it is like in the relevant sense, does not require there being something *for the subject.* This part will support the division of the research for a theory of consciousness into a theory of the qualitative character on the one hand and a theory of the subjective character (for-meness) on the other.

For-meness is required for distinguishing a subjective use from a non-mental use of 'what it is like'; we are looking for a theory of *subjective* qualities. There is something it is like to be the table I am writing on; namely, being made of wood, painted in black, having four legs, etc. However, this is not the use of 'what it is like' we are interested in; we intend to capture a subjective use, something completely different, phenomenal consciousness as 'what it is like to undergo an experience'. In this second sense, there is nothing it is like *for the table* to have me writing on it (or so I think). That has been noted by Rosenthal (2005):

> It is important to distinguish this somewhat special use of the phrase what it is like to describe subjectivity from its more general, non mental use. There is something it is like to be a table, or even to be this very table. What it is like to be a table, for example, is roughly something having characteristic features of tables. But this is of course not what's involved in talking about what it's like to have an experience. (Rosenthal, 2005, p. 656)

Thomas Nagel (Nagel, 2002), who famously introduced the phrase, made it clear that phenomenal consciousness requires there being something for the subject:

> [T]he fact that an organism has conscious experiences at all means, basically, that there is something it is like to be that organism... But fundamentally an organism has conscious mental states if and only if there is something it is like to *be* that organism –something it is like *for* the organism. (Nagel, 2002, p. 219)

One interesting way of presenting the subjective character is appealing to what is phenomenologically shared by all phenomenally conscious states. To a first approximation, the best way to point out to this common element is, I think, by examples. You can distinguish between experiences as of different shades of red, say $\mathtt{red}_{35}$ and $\mathtt{red}_{40}$. These two experiences are more similar, phenomenologically speaking, between them that with regard to an experience as of $\mathtt{red}_2$. Furthermore, experiences as of $\mathtt{red}_{35}$, as of $\mathtt{red}_{40}$, and as of $\mathtt{red}_2$, seems to be more similar between them that an experience as of $\mathtt{green}_3$. In general we distinguish between experiences as of red from experiences as of green.

The phenomenal character of experiences as of red and experiences as of green are in a sense different. But they are also in a sense similar (the similarities and differences are here meant to be phenomenological): they are color experiences. They differ, in a sense, from visual experiences of forms, like a visual experience as of a square. And again, these experiences have something in common, they are all visual experiences, and in a sense the *way* they *feel*, their phenomenal character, is similar.

Similarly, auditory experiences of an A produced by a violin are more similar to those produced by a viola than those produced by an electric guitar. The experience of an A played by a violin, and the experience of an A one octave below by the same violin have something in common and all the experiences of the notes of a violin have something in common. All auditory experiences have phenomenologically something in common. Tactile experiences have something in common, the same for auditory experiences, visual experiences, taste experiences, pains, orgasms, etc; and all experiences have something phenomenological in common. They are, so to say, marked as my experiences. Phenomenally conscious experiences happen *for the experiencing subject* in an immediate way and as part of this immediacy they are implicitly marked as *my* experience. This is what I call the subjective character of the experience. All these phenomenally conscious experiences have something in common, their distinct first-personal character. All phenomenally conscious experiences have this quality of for-meness or me-ishness.

Someone could suggest at this point that the subjective character, as I am presenting it, is simply another quality of my experience. If this is a claim about the name it deserves, I still prefer to keep a different name to mark that whereas different kinds of phenomenally conscious experiences have different qualitative character, all phenomenally conscious experiences share a subjective character. Gallagher and Zahavi (2006) nicely present this idea:

> The mineness in question is not a quality like being scarlet, sour or soft. It doesn't refer to a specific experiential content, to a specific what; nor does it refer to the diachronic or synchronic sum of such content, or to some other relation that might obtain between the contents in question. Rather, it refers to the distinct givenness or the how it feels of experience. It refers to the first-personal presence or character

> of experience. It refers to the fact that the experiences I am living through are given differently (but not necessarily better) to me than to anybody else. It could consequently be claimed that anybody who denies the for-me-ness of experience simply fails to recognize an essential constitutive aspect of experience. Such a denial would be tantamount to a denial of the first-person perspective.

The subjective character points to some form of intimate relation between the subject and her conscious experiences. The first thing that should be noted is a phenomenological observation: whenever a subject has an experience, certain quality is somehow given to her, there is a special relation between the subject and the experience. Some philosophers have maintained that the subjective character can be characterized as a certain form of awareness. For instance, Kriegel, who has carefully developed the distinction between qualitative and subjective character (Kriegel (2005, 2006, 2009)), presents the subjective character as follows:

> We may construe phenomenal character as the compresence of qualitative character and subjective character.
>
> To say that my experience has a bluish qualitative character is to attribute to my experience the property of exhibiting a certain specific sensuous quality. It is not to say that the property in question is irreducible, or intrinsic, or inexplicable. It is merely to assert the existence of that property.
>
> To say that my experience has a subjective character is to say that I am somehow aware of my experience. Conscious experiences are not sub-personal states, which somehow take place in us and which we "host" in an impersonal sort of way, without being aware of them. Mental states we are completely unaware of are unconscious states. So when I have my conscious experience of the sky, I must be aware of having it. In this sense, my experience does not just take place in me, it is also for me. Again, by asserting the existence of the property of subjective character, I do not mean to imply that it is irreducible. (Kriegel, 2006, p. 199)

Kriegel points toward a certain form of awareness as characteristic of the subjective character. Being in a phenomenally conscious mental state feels some way or other. There makes no sense to talk about a feel we are completely unaware of; mental states I am completely unaware of are not conscious mental states at all. Awareness seems to be a certain form of access. Some philosophers motivated by this idea have criticized Block's distinction between access and phenomenal consciousness (e.g. Rosenthal (2005)). But the distinction between qualitative and subjective character as constitutive parts of the phenomenal character is perfectly compatible with there being a distinction between access and phenomenal consciousness. The question lies on the kind of awareness that is essential to subjective character. Block himself concedes the possibility of some form of awareness being constitutive of phenomenal consciousness.

> We may suppose that it is platitudinous that when one has a phenomenally conscious experience, one is in some way aware of having it. Let us call the fact stated by this claim – without committing ourselves on what exactly that

> fact is – the fact that phenomenal consciousness requires Awareness. (This is awareness in a special sense, so in this section I am capitalizing the term.) Sometimes people say Awareness is a matter of having a state whose content is in some sense "presented" to the self or having a state that is "for me" or that comes with a sense of ownership or that has "me-ishness" (Block, 2007a, p. 484)

I want to leave open at this point the characterization of Awareness, as Block calls it. In chapter 5, where I address the problem of the subjective character of the experience, I will deal with this issue in detail.

1.3.3 *Phenomenal Character and the Problem of Consciousness*

Though not everyone would agree with the distinction between qualitative character and subjective character, this distinction is, to the very least, useful for making a taxonomy of philosophical theories of consciousness. A simple look into the literature about consciousness reveals, as Kriegel (2009) has noted, that different theories about phenomenal consciousness, broadly understood as the property responsible for the mystery of consciousness, seem to target different phenomena.

On the one hand, there are theories that maintain that the root of the problem of phenomenal consciousness is the qualitative character. The qualitative features that compose the qualitative character of the experiences are usually referred as 'qualia'.[17] Representational theories of consciousness are a characteristic example of this position (Dretske (1993); Kirk (1996); Tye (1997, 2002)). On the other hand, the so called higher-order theories of consciousness target the subjective character as the property where the mystery lies in (Armstrong (1981); Carruthers (2000); Lycan (1996); Rosenthal (1997, 2005)).

There are two questions that should be attended to: what is the relation between phenomenal character on the one hand and qualitative and subjective character on the other? And what is the relation between qualitative and subjective characters? Different theories offer different answers.[18]

With regard to the second question we can distinguish four different positions:

1. The qualitative and the subjective character can both be instantiated independently.
2. The qualitative character is a constitutive part of the subjective character.
3. The subjective character is a constitutive part of the qualitative character.
4. Neither qualitative character nor subjective character can be instantiated independently of each other.[19]

These four positions are intended to be mutually exclusive. For instance, if someone believes in 2, then she does not believe 1 (3 or 4 either),

17 I prefer to continue the discussion in terms of character or qualitative property instead of qualia. I find use of the term 'qualia' in the literature confusing and different philosophers seem to refer to different properties by 'qualia'. I hope the notion is clear enough for the moment. I will say more on the qualitative character in chapter 4.

18 In this taxonomy I am following the one presented by Kriegel (2009, pp. 52-53).

19 The kind of possibility involved in 1 and 4 is metaphysical possibility.

because in this case the qualitative character could be instantiated independently of the subjective character but not the other way around.

In a first step I have introduced phenomenal character as the property of the subject's experience responsible for the problem of consciousness. So understood, the question on the relation between phenomenal character on the one hand and qualitative character and subjective character on the other may receive different answers. We can distinguish the following views:

QUALITATIVISM Phenomenal character is identical with qualitative character.

SUBJECTIVISM Phenomenal character is identical with subjective character.

COMPRESENTISM Phenomenal character is identical with certain combination of phenomenal character and phenomenal character.

The preferred theory in the first distinction combines with the one in the second distinction. For instance, there can be qualitivists that maintain that qualitative character is separable from subjective character and qualitivists that deny it, similarly for the subjectivist. We can consider the following possible combinations of views about phenomenal consciousness:

SQ Separatist Qualitativism: combines Qualitativism with either 1 or 2.

SQ maintains that phenomenal character is identical to qualitative character which is separable from subjective character. In the same way we can define other possible alternatives:

IQ Inseparatist Qualitativism: combines Qualitativism with either 4 or 3.

SS Separatist Subjectivism: combines Subjectivism with 1 or 3.

IS Inseparatist Subjectivism: combines Subjectivism with either 4 or 2.

SC Separatist Compresentism: combines Compresentism with 1.

IC Inseparatist Compresentism: combines Compresentism with either 2, 3 or 4.

I find separatist views unappealing (SS, SI and SC) for different reasons.

In the first place, SC seems to be committed to the view that phenomenal character is beyond the mere addition of qualitative and phenomenal character. If there can be mental states of a subject that instantiate qualitative properties, and similarly for the subjective character, without thereby S having a phenomenally conscious experience, then phenomenal character has to be something beyond the mere addition of these properties. Phenomenal properties would be some kind of emergent properties essentially different from both qualitative and subjective properties because only when these two are combined the mystery of consciousness arises.

My reason for rejecting SQ is simply that this position maintains that subjective character plays no constitutive role at all in phenomenal character, what I find implausible. As I have argued above, phenomenal character necessitates subjective character.

According to SS, phenomenal character is identical to subjective character which is separable from qualitative character. The idea of a phenomenally conscious experience without a quality doesn't seem to make sense and I do not know of any theory that maintains something in the vicinity.

I find inseparatist theories more appealing.

IQ is an interesting position. One may read Tye's PANIC (Tye, 1997, 2002) theory as an example of IQ theory. In section 5.2, I will present some arguments for not endorsing such a view.

I am more sympathetic to IS or IC. I think that the subjective character lies at the heart of the mystery, I find it more puzzling. As a materialist it seems mysterious that, if mental states are just physical states, it can be that they *'feel'* at all (the problem of the *way* they *feel* is the problem of the qualitative character). I am puzzled about something common to all phenomenally conscious states, the subjective character. Higher-order theories, for instance, are constructed as inseparativist subjectivist theories. However, I think that we have good reasons for preferring a first-order approach to the problem of consciousness as I will argue in 5. One very interesting example of first-order IS theory is Kriegel's self-representationalism, however I think that it faces serious objections.

The view I am going to present and defend in this dissertation is a form of inseparatist compresentism, where the subjective character is constitutive of the qualitative character. So, it combines compresentism with 3. I will call this theory Self-Involving Representationalism (SIR).

1.4 THE STRUCTURE OF THE DISSERTATION

This dissertation is organized in two parts besides this introduction: 'Consciousness and Materialism' and 'A naturalist theory of consciousness: SIR'.

The first part presents some of the problems that a materialist theory of consciousness has to face; it presents and tries to rejoin arguments against materialist theories of consciousness. It is organized in two chapters.

The purpose of the first one (2), Consciousness and Materialism, is to rehearse the two main arguments presented in the last thirty years against materialism: the modal argument and the knowledge argument. I am going to present these interrelated arguments and discuss the plausibility of some possible replies to counter these arguments. I will argue that there is a reply to these arguments that is compatible with the thesis of materialism: the phenomenal concept strategy.

According to the the modal argument, what is conceivable *in the right way* is metaphysically possible, at least in the case of phenomenal consciousness. If zombies are conceivable *in the right way* then they are metaphysically possible and materialism is false. I will characterize this *right way* and deny that the metaphysical possibility of zombies is entailed from this kind of conceivability. The remaining work for the materialist is to explain the conceivability of zombies. This work is done by the phenomenal concept strategy.

The knowledge argument also exploits the lack of an a priori entailment between physical truths and phenomenal truths to show a problem in the explanation of the nature of phenomenal consciousness

that jeopardizes materialism. I will accept that there is an explanatory gap, but deny that this gap has metaphysical consequences.

The lack of a priori entailment that grounds the explanatory gap and the conceivability of zombies are both explained in terms compatible with materialism by the so called phenomenal concept strategy. According to the phenomenal concept strategy, the anti-materialist arguments take their force from a misunderstanding of the special nature of phenomenal concepts, the concepts we deploy for referring to phenomenally conscious experiences. I will first introduce the strategy and then defend it against some arguments that have been presented.

The second chapter of this part (chapter 3), Phenomenal Consciousness and Vagueness, deals with arguments that maintain either that phenomenal characters are vague and physical properties are not or that phenomenal characters are sharp and physical properties are vague. From this they conclude that phenomenal characters cannot be identified with physical properties.

I will briefly introduce the phenomenon of vagueness. Then I will first consider arguments that maintain that phenomenal characters are vague whereas physical properties are not. This would jeopardize some naturalistic theories if the arguments are sound. To clarify the debate, I will distinguish two senses in which phenomenal characters can be said to be vague: horizontally and vertically. The first one is related to the qualitative character of the experience, the latter to the subjective character.

The non-transitivity of the relation 'looks the same as' has been used to support the claim that phenomenal characters are horizontally vague. I will argue that this mistakes the notion of distinguishability that should individuate phenomenal characters and that it presupposes that cognitive access is essential to the phenomenal character. I will further consider arguments that support the claim that phenomenal characters are vertically vague; namely, that it can be indeterminate whether being in a state feels at all or not. I will maintain that these arguments are based either on a confusion on the notion of consciousness in play or on a confusion between metaphysics and epistemology. Finally, I will consider an argument that accepts that phenomenal characters are sharp but not so physical properties and argue that this is not a problem for materialism.

In the second part of this dissertation I present the mainstays of a naturalistic theory of consciousness, the SIR theory. This part is also divided into two chapters each one devoted respectively to one aspect of phenomenal character: the qualitative character and the subjective character.

The 4th chapter is about the qualitative character of experience. Some theories maintain that qualitative characters are extrinsic properties of the subject: if this is right, then microphysical duplicates may not be phenomenologically identical. I want to hold on to the intuition, supported by our current knowledge of the brain, that phenomenal characters are intrinsic properties of the subject. I will start this chapter by reviewing some of the theories that reject this claim. I will briefly introduce direct realism and the well-known problem of hallucinations. Representationalism solves the problems of direct realism that arise with hallucinations by appealing to the relation of representation: the content of mental states, what mental states represent, is the same in cases of veridical perception and hallucination. The content of the

experience determines the qualitative character of the experiences; i.e., qualitative properties are representational properties. The differences between two phenomenally conscious experiences are differences in the property represented, differences in the content of the experience. One of the most attractive reasons to embrace representationalism is the transparency of experience: to a first approximation, the transparency of experience shows that when we introspect the phenomenal character of the experiences we look "through" phenomenal properties and all that we do is to focus on the properties of the perceived object.

I will discuss some objections to the representationalist view. Representationalism, I will argue, has resources to deal with these objections. One of the objections I will present, the shifted spectrum objection, is especially pressing for those forms of representationalism that hold that representational properties are extrinsic properties of the subject. I will argue that narrow representationalism, the brand of representationalism I will embrace, can address this objection. According to narrow representationalism, the content of the experience that determines the phenomenal character supervenes on the intrinsic properties of the subject and so qualitative properties are intrinsic properties of the subject.

There are two questions that require further consideration:

1. What is the content of phenomenally conscious experiences such that it supervenes on the intrinsic properties of the subject?

2. In virtue of what does the relation of representation between that which is represented (the content) and that which does the representing (the vehicle of representation) hold?

I will provide a characterization of representational properties that respects the intuition that phenomenal properties are intrinsic properties of the subject; this characterization should also address the problem of the shifted spectrum presented in the previous section. I will argue that the content of the experience is *de se* content: the content of an experience with phenomenal character PC is the *centered feature* of having the disposition to cause that experience in me *in normal circumstances*. I will explain and justify the notion of centered feature and I will dispel any worries about circularity that this very rough description of the view may bring about.

Naturalistically oriented theories of mental content appeal to the notion of function to explain the representation relation: the content of a mental state is what the mental state has the function of indicating. I will explore several of these theories of function. I distinguish between etiological and non-etiological theories of function. The former maintain, whereas the latter deny, that the function of a trait depends on its causal history. I will argue that etiological theories cannot be a satisfactory option.

The last chapter (5) presents in detail the notion of subjective character. I will argue that all my phenomenally conscious experiences have something in common; a common first-person perspective in which a certain quality is given to me. I will offer two arguments in favor of the subjective character of phenomenally conscious experiences. The first is based on a phenomenological observation and the second, for those who are skeptical about the phenomenological observation, based on the analysis of the content of experience presented in the previous chapter.

I will argue first against theories that try to explain the subjective character of experience as some form or other of cognitive access and discuss two arguments that suggest that a mental state can be phenomenally conscious without being accessed by any cognitive process. Then I will present theories of consciousness that explain the subjective character of the experience as a representational relation. According to these theories, phenomenally conscious mental states are mental states that are *adequately represented*. As we will see, different theories provide different characterizations of what *being adequately represented* means.

Representational theories of the subjective character can be divided into two groups depending on whether the mental state is represented by a numerically distinct mental state (higher-order) or not (same-order). I will present several arguments to expose some problems that these theories face and my reasons for rejecting them as a plausible account of the subjective character of experience.

In the last section, I present my own proposal that I hope satisfactorily accounts for the subjective character of experience while avoiding the problems faced by other theories.

Part II

CONSCIOUSNESS AND MATERIALISM

2 CONSCIOUSNESS AND MATERIALISM

Materialists are committed to provide a satisfactory reply to the following question: how can we explain, in physical terms, what makes it the case that undergoing an experience *feels*? Or, at least, they have to explain how it can be possible that phenomenal consciousness is not explainable in physical terms, while being indeed physical in nature.

Some philosophers have presented arguments to the effect that any attempt to explain consciousness in physical terms is condemned to fail and that, despite the good reasons we have for believing in it, materialism is false.

The purpose of this chapter is to rehearse the two main arguments presented in the last thirty years against materialism: the modal argument and the knowledge argument. I am going to present these interrelated arguments and discuss the plausibility of some possible replies to counter these arguments. I will argue that there is a reply to these arguments that is compatible with the thesis of materialism: the phenomenal concept strategy.

In section 2.1, I will start by introducing the modal argument as presented by Kripke. Functionalism, the proposal I will be favoring in the following chapters, seems to be an acceptable position immune to this argument. With this excuse, I will introduce in 2.1.2 functionalism and discuss three different concerns for the functionalist approach. The first one deals with its plausibility, the second one with an epistemic question, and the third one with its compatibility with materialism.

Chalmers has presented a refined version of the modal argument that also jeopardizes functionalism. This refined version is presented in 2.1.3. Zombies, microphysical (and functional) duplicates of us lacking phenomenal consciousness, seem to be conceivable. If there is an entailment between some form of conceivability and metaphysical possibility and zombies are conceivable in this sense, then materialism is in trouble.

I will argue in 2.1.4, following Balog, that we have a good reason for rejecting the argument as unsound. The reason is that accepting the very same principles that back up the premises of Chalmers' version of the modal argument we can derive the conclusion that materialism is false a priori –a conclusion that the materialist can easily reject. The remaining work for the materialist is to explain the conceivability of zombies. This work is done by the phenomenal concept strategy.

In section 2.2 I will present the knowledge argument. This argument exploits the lack of an a priori entailment between phenomenal truths and physical truths to show a problem in the explanation of phenomenal consciousness that jeopardizes materialism. There is an explanatory gap between phenomenal truths and physical truths.

In subsection 2.2.1, I will make some considerations about the knowledge argument. Some philosophers have claimed that reductive materialism seems not to be in a worse position than reductive dualism. Surely, if we are looking for a satisfactory theory of phenomenal consciousness this reply is not enough: we still need to explain, in a way compatible with materialism, the reasons of a failure in the a priori entailment of phenomenal truths from physical truths. Second, I will consider some

remarks made by Brown, which cast some doubts on the knowledge argument. I will maintain that Brown's argument is appealing but insufficient for rejecting that there is a problem for materialist theories.

Advocates of the explanatory gap hold the following thesis:

> (A priori entailment thesis)
>
> If P reductively explains H, then it is a priori that $P \rightarrow H$.

In subsection 2.2.2 I will discuss the a priori entailment thesis in some detail and three different views on the relation between a priori entailment and reductive explanation. I will conclude that the lack of a priori entailment shows that there is a failure in the explanation exclusive of phenomenal consciousness (or at least not ubiquitous in scientific reductive explanations) and also deny that the right conclusion to be derived from this gap is the truth of dualism. To block this conclusion, an explanation of the failure in the a priori entailment has to be provided. This is the task of the phenomenal concept strategy presented in the last section.

Section 2.3 presents and defends the phenomenal concept strategy. According to the phenomenal concept strategy, the anti-materialist arguments take their force from a misunderstanding of the special nature of phenomenal concepts, the concepts we deploy for referring to the phenomenal character of our experiences. I will first present the strategy and then defend it from two arguments: the first one maintains that phenomenal concepts are not special at all; the second one holds that either phenomenal concepts cannot be explained in a way that is compatible with materialism or, if they can, then what cannot be explained is our epistemic situation with regard to the gap. In either case materialism would be jeopardized.

I will offer a rejoinder to these arguments and conclude that the phenomenal concept strategy offers a satisfactory reply, compatible with the truth of materialism, to the modal and knowledge arguments. So, materialism has nothing to fear from these arguments.

2.1 THE MODAL ARGUMENT

2.1.1 *Kripke's modal Argument*

In the lectures 'Naming and Necessity', Saul Kripke (1980) introduced the notion of *rigid designator* in the course of his argument against descriptivism. Rigid designators are those terms that pick out the same referent in all possible worlds in which that referent exists and do not designate anything else in those possible worlds where the referent does not exist. Proper names are paradigmatic examples of rigid designators. The term 'Sebas' will pick me out, the writer of this dissertation, in every possible world in which I exist, including those worlds in which I decided not to study philosophy, and therefore I do not write this dissertation. 'Sebas' will pick out no one in those possible worlds in which I do not exist. Kripke argues that proper names and certain *natural kind terms* designate rigidly.

Moreover, he maintains that identity statements involving two rigid designators are, if true, necessarily so -they are true in every possible world where the terms involved in the identity statement refer. Some of these identity statements are *a posteriori*, such identities are justified on

the basis of empirical knowledge and unknowable by mere reflection on the concepts involved.

With this technical tool in hand, Kripke presents an argument (the modal argument), that goes back to that of Descartes for the distinctness of mind and body, against a reduction of phenomenal properties to physical properties. In particular, Kripke argues against psychophysical identities, that is, the identification of the referent of a scientific term with the referent of a phenomenological term. Examples of scientific terms are 'H_2O', 'motion of the particles', 'the chemical element with atomic number 79', 'C-fiber stimulation', etc. Examples of phenomenological terms are those that refer to *feelings*, to experiences with certain phenomenal character: 'sensation of pain', 'sensation as of red', etc.

Some scientific identities identify the referent of ordinary folk terms with the referent of scientific terms, like 'water is H_2O' or 'Heat is the motion of particles'. These identities are, if true, necessary and *a posteriori*. They are *a posteriori* because no matter how ideally we would reflect on the concept 'water', we could not discover the nature of its reference by a priori reflection, we need of empirical research to come to know such essences. They are, however, necessary, or so argues Kripke, because the terms involved are rigid designators. If they pick out the same referent in the actual world, then the identity statement is true, and given that they pick the same entity in every possible world, the statement is true in every possible world; i.e., necessary.

Contrary to *a priori* necessities, such as 'two plus two is four' or 'no bachelor is married', *a posteriori* necessities seem to be contingent being nevertheless necessary. There is a sense in which it seems to us that water could have turned out to be a different substance than H2O. On the other hand, a competent speaker would not consider that there can be a bachelor who is married. How do we account for this appearance of contingency in *a posteriori* necessities?

Kripke himself offers a way to explain the appearance of contingency away, one that is not available in the case of psychophysical identities.

Consider the two examples presented by Kripke:

(a) Pain is C-fibers firing.[1]

(b) Water is H_2O.

In both examples we are presented with identity statements involving rigid designators. Following Kripke, both, if true, must be necessarily true.

There is an appearance of contingency in both cases; i.e., it seems conceivable that they are false. However, if they are true, they are necessarily true and that means that there is no possible world where the terms refer and the identity does not hold.

We need to explain the apparent contingency away -that is, we need to explain why there is an illusion of contingency. According to Kripke this is possible for (b) but not for (a).

Kripke suggests that when someone considers that water could have turned out not to be H_2O, what she is actually considering is an imagined scenario where there could have been another entity, W, with the manifested properties of water (being colorless, odorless, filling

1 Nowadays, no one would maintain that pain is C-fibers firing. The reader is free to replace 'C-fibers firing' for her favorite neural correlate of pain. By neural correlate of an experience with phenomenal character Q, I mean the brain activity that is minimally sufficient for having an experience with phenomenal character Q.

up lakes, etc.), which is not H_2O. Kripke denies that W is water. It is true that the referent of 'water' (i.e. water) is fixed by some contingent properties like being colorless, odorless, etc. and that in the imagined scenario those very same contingent properties could be used to fix the referent of 'water', but it doesn't follow from that that water is W and that the imagined scenario is one in which water is not H_2O. Because, even if we use some contingent properties to fix the referent of a term, those properties do not determine its meaning and it doesn't follow that our term will refer to anything that has those properties in any possible world (remember that the properties are contingent).

Moreover, as Kripke claims, given that 'water' and 'H_2O' are rigid designators, they have to refer to the same substance in any possible world if they, in fact, refer to the same substance in the actual world. Therefore, if W is not H_2O, then 'water' (as used in the actual world) does not refer to W, an W is not water. So, briefly, the imagined scenario is one in which some colorless, odorless, etc. liquid is not H_2O and therefore is not water. W fulfills the same role as water does, but not being H_2O, it is not water.

Unfortunately for the materialist, this way of explaining the apparent contingency of a posteriori and necessary identity statements is not available for (a). The properties that help fixing the referent of phenomenological terms are essential to that referent. For example, we individuate pain by its manifested properties, by the way it feels: nothing could be felt like pain and fail to be pain.

Imagine an alien creature called Kodos. Kodos lacks C-fibers and neurons; it has a completely different cognitive system. Like Lewisian martians (Lewis 1978), Kodos has a brain consisting of fluid and inflatable cavities. A knock causes inflation of small cavities in its feet and when these cavities are inflated Kodos is in pain. When its friend Kang hits it, Kodos tries to avoid a second knock, moves away from Kang and yells complaining. Imagine also that there is something it is like for Kodos to be hit by Kang, and that the phenomenal character of Kodos experience is like the one I have when I am in pain. We would say that Kodos is in fact feeling pain: the properties used to fix the referent of 'pain' do determine also its meaning and, therefore, anything that has these properties in any world necesarily will be pain. That is why we cannot provide the same explanation as in (b) and claim that what we are considering is another phenomenon that feels like pain but is not pain, for everything that feels like pain is pain.

Kripke claims that that the identity (a) seems to be contingent whereas it has to be necessary if true. We cannot explain away the appearance of contingency in identities like (a). Identities involving rigid designators are if true necessarily so; i.e. they cannot be contingent. A bit more formally, Kripke's argument is the following:

Let 'P' be a term that rigidly refers to a physical property

Let 'Q' be a term that rigidly refers to a phenomenal property

(Kripke's Argument)

(A1) P=Q assumption

(2) Identities involving rigid designators are necessary, if true.

(3) P=Q seems contingent. It seems that we can conceive that P≠Q, given that such an indentity is a posteriori.

(4) The only explanation that we have that is compatible with (1), (2) and (3) is that an a posteriori necessary truth seems contingent because one of the referent of the terms is picked out through a contingent property.

(5) Q's referent is not picked through a contingent property.[2]

(6) It is not necessary that P=Q (From 3 to 5).

∴ P=Q is false (From 1 to 6 by reductio ab absurdum).

Kripke's argument is a *reductio* of the materialist thesis. The materialist has several options to reply to this argument. Functionalism is one of them. Functionalism is not committed, as we are about to see, to psycho-physical identities, but to conditionals of the form $P \rightarrow Q$.

We will see in 2.1.2 that the modal argument can be refined in such a way that it targets also functionalism and therefore further tools are required to rejoin the anti-materialist objections. I will present these tools along the chapter and in particular in 2.3. However, given that the proposal I will make in the second part of this dissertation is a functionalist one and that it seems to be a reply to the modal argument as presented by Kripke, I will introduce the main insights of functionalism during the next subsection.

2.1.2 *Functionalism and Materialism*

Many materialists are not committed to identities such as (a). They try to explain phenomenal consciousness in terms of postulated mental functions, the idea being to identify phenomenal properties with certain functional properties. Let me start by presenting the notion of function in more detail.

According to functionalism, some systems have a functional organization. This organization is individuated by three elements:

1. The number of abstract elements in which the system can be decomposed.
2. The number of possible states for every element (in the most simple case will be on/off).
3. The relation between a state of a component and the rest of the states of all the elements of the system, and how the output of the system and the transitions from one state to the next depend on the previous states and inputs.

A finite state machine[3] will be useful for illustrating an example. Consider a dispensing machine. The machine sells beer by the bottle at 1€,

2 I am assuming here, as Kripke does, that P is not picked out through a contingent property. For a reply to Kripke's argument the rejects this premise maintaining that P picks out its referent through a contingent property see Boyd (1980).

3 The example is a Mealy machine, a Finite State Machine (FSM) whose output values are determined by both the current state and the input of the system. A FSM is a mathematical abstraction used in digital design, the one in the example can easily be used for the design of a beer expending machine.
Though philosophers commonly appeal to Turing machines for computational abstractions I consider it more illuminating to present the FSM in the figure. Turing machines are not very useful in real design. If one is familiar with Turing machines, a FSM can be seen as a Turing machine where the ability to rewrite the tape has been removed

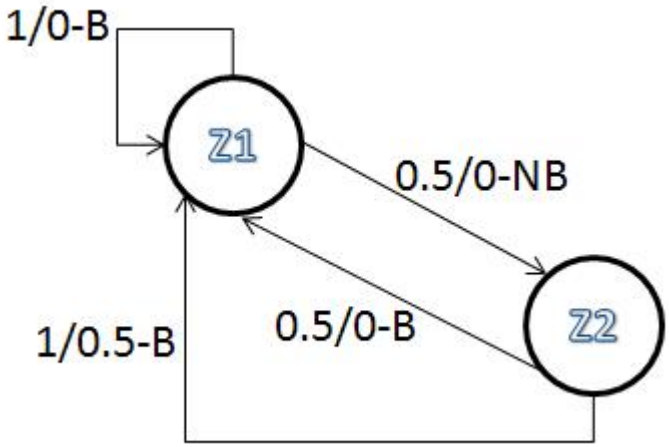

Figure 2: Example of a Finite State Machine.
The finite state machine that describes the function of a beer dispensing machine has two states Z1 and Z2. One input with two possible values, either 1€ or 50 cents (represented as the values near the arrows before the slash by 1 or 0.5) and two outputs, one for the return and one for the beer (represented after the slash by 0 or 0.5 in the case of the return, and B or NB representing whether the dispensing machine provides a bottle or not)

and accepts 1€ and 50 cent coins. The state machine is represented in figure 2.

There is only one element in the system, the selling machine, with two possible states: Z1 and Z2. Z1 is the state the machine is in when waiting for a coin. If a 1€ coin is introduced it dispenses a bottle. If a 50 cents coin is introduced, then state Z2 is activated and no bottle is dispensed. When the machine is in Z2, it is waiting for a 50 cents coin; if such a coin is introduced, the machine goes to Z1 and provides a bottle. If, while being in Z2, a 1€ coin is introduced, it goes to Z1 and returns 50 cents in change and a bottle.

This simple machine is multiply realizable. It can be made out of wood, plastic or iron; it can be implemented with transistors, vacuum tubes or by myself being inside a box taking the money and dispensing the beer. The function will be the same. If satisfying this function is all that is relevant for being a beer dispensing machine, all the former realizations count as beer dispensing machines.

Barker-Plummer (2011). FSM are less powerful than Turing machines, since they cannot use the tape to remember the state of the computation.
A Mealy machine is, formally, a 6-tuple, $(S, S_0, \Delta, \Omega, T, O)$, where:

s is a finite set of states

s_0 is the initial state, an element of S

Δ is the set of inputs

Ω is the set of outputs

t is the transition function ($T : S \times \Delta \to S$) that maps pairs of a state and an input symbol to the corresponding next state.

o is the output function ($O : S \times \Delta \to \Omega$) that maps pairs of a state and an input symbol to the corresponding output symbol.

Alternatively the same functional system can be described by a Moore machine. A Moore machine is a FSM whose output values are determined solely by its current state.

With regard to the mind, functionalists claim that the proper characterization of mental states is given by their functional role within a system.[4] Phenomenally conscious states are the states that occupy a certain role in such a system, like Z1 or Z2. Of course, our mental life is much more complicated than the presented state machine, but the main idea remains: the individuation conditions of a certain mental state are given by a certain functional role; by the role the state has within the organism.

In this sense, a certain mental state is phenomenally conscious not in virtue of its internal constitution; i.e. being a certain neural network firing at a certain frequency, but in virtue of the role it plays in the system. For instance, phenomenal consciousness satisfies certain function in relation to other aspects of cognition. Functionalists do not identify phenomenal properties with certain physical properties but with certain functional role within a system. The corresponding functional role can be implemented in different ways.

According to functionalism, mental states are multiply realizable. If we buy into functionalism, the intuition that the alien lacking C-fibers could be in pain is explained in terms of the multiple realizability of mental states. To be in pain is to be in a certain functional state, a state shared by Kodos and me, that is implemented in a different way in each of us.

The first problem for functionalism arises as the question about the level of abstraction required for selecting the elements: how finely do we need to individuate the parts of the system in order to get the function of phenomenal consciousness? If the brain were the supervenience base of phenomenal consciousness, the resulting functionalization will be very different if we take hemispheres, lobes, neurons or molecules to be the relevant parts. The level of abstraction will depend on the theory of phenomenal consciousness.

One possible candidate for identification with phenomenally conscious mental states are states within our folk psychology. In this case, the theory on which the level of abstraction for individuating states will depend is our folk psychology. A phenomenally conscious mental state is a state that plays a certain functional role within our folk psychology. For instance, in the case of pain, the mental state that tends to be caused by body injury, tends to trigger the belief that something is wrong with the body and the desire to be out of that state, etc.[5]

This description seems unsatisfactory. There is more to being in pain than the description in folk psychological terms given above, and this 'more' is precisely what we are interested in, if we are interested in phenomenal consciousness: being in pain *feels* a certain way. None of those functional aspects seem to be essential to pain, to the way it *feels*, to the phenomenal character of an experience as of pain. More interesting levels of description arise from moving toward psychofunctionalism (Block and Fodor 1972; Block 1978) or neurofunctionalism for instance, where the theory that is relevant for defining the roles is that of empirical psychology or neuropsychology.

4 Different theories consider different systems as we will see in chapter 4.

5 Formally, mental states and processes are treated as being implicitly defined by the Ramsey sentence (Lewis 1972) of our folk psychological theory, free of any mental state term. If THEORY is our folk theory with m mental states of which 'pain' is the nth term, where S is the set of states, Δ the set of inputs and Ω the set of outputs, it is possible to define *X is in pain* as:
X is in pain=X is such that $\exists S_1, \exists S_2, ..., \exists S_m [\mathrm{THEORY}(S, \Delta, \Omega) \& x \text{ is in } S_n]$

I am not going to argue in favor of any theory at this point; I will deal more specifically with theoretical and empirical considerations for functionalization and identity of phenomenally conscious mental states in chapters 4 and 5.

In the remaining of this section I will discuss three different very general concerns about the functionalist approach that bear on phenomenal consciousness; these concerns do not depend on the particular form of functionalism that is preferred. The first one deals with its plausibility, the second with an epistemic question, and the third one with its compatibility with materialism.

The Plausibility of a Functionalist Approach

Functionalist approaches to phenomenal consciousness have been famously criticized by Ned Block. In (Block, 1978), he presents several thought experiments to motivate the claim that functionalist accounts of phenomenal consciousness are implausible.

The first thought experiment is known as *inverted spectrum*. The idea of an inverted spectrum has its origins in Locke (1994) and has been discussed by other philosophers like Wittgenstein (1968). The idea is the following: there could be a subject with an *inverted spectrum* (someone who has an experience as of red when looking at the grass, as of green when looking at a ripe tomato, as of yellow when looking at the sky, etc.) who is behaviorally indistinguishable from someone with normal color vision.[6] Similarly, Block argues that there could be a subject S being in an state satisfying the functional description of the state I am in when I undergo an experience as of red such that S is having an experience as of green instead. If the mental states of two subjects have the same functional role but their experiences differ in their phenomenal character, then phenomenal properties cannot be identified with functional properties.

For this objection to succeed, spectral inversions that are not behaviourally detectable would have to be metaphysically possible, and the asymmetries in color space are a good reason for thinking that they are not. Given that there are more perceptually distinguishable color shades between red and blue than there are between green and yellow, a red-green inversion would be behaviourally detectable (Hardin 1997). Kalderon and Hilbert (2000) further argue that "every possible quality space must be asymmetrical" (ibid., p. 204) and so, inverted scenarios are not possible.

Another thought experiment is known as the *absent qualia*. The idea of this second argument is that there could be a system that is functionally equivalent to a human being but which nevertheless undergoes no phenomenally conscious experience. In the *Chinese Nation* thought experiment (Block 1978) the Chinese Government wants to generate a human mind. For that purpose, they study Block's brain. They come to know the activity of every single neuron in his brain, particularly when he is feeling an intense headache. They recruit the population of China to duplicate Block's neurons (at the time when the experiment takes place, the population of China is greater than the number of neurons in a brain, over 100 billions), with each Chinese volunteer instructed to simulate one neuron.[7] Let's concede to Block that the

6 See also Shoemaker (1982)

7 The intuition that Block is trying to put forward is, I think, independent of the theory used for ramseyfication (as we saw in footnote 5). In the original example, each person is

technology at the moment of the experiment and the hard training of the participants allow them to carry out their task. Some of them have to communicate with many colleagues, since certain neurons have more than 60.000 connections. The participants have satellite transmitters for communicating with each other, thereby simulating the synapses between neurons. They receive the equivalent of sensory input from an artificial body and send instructions back to the body for acting in the world. Block claims that, intuitively, contrary to what happens to him, this system does not feel any pain, it has no phenomenally conscious experience at all. Given that the system is a functional duplicate of him and he has phenomenally conscious experiences, if the *Chinese Nation* lacks them, then phenomenal consciousness cannot be a matter of functional organization.

For many readers, the *Chinese Nation* is not the kind of thing we would ascribe phenomenally conscious experiences to. Intuitively, the *Chinese Nation* is not the kind of thing that gives rise to phenomenally conscious experiences.

Block's argument has a certain intuitive force, but this is far from being a knockout argument, especially if we consider, as Chalmers (1996, p. 235) notes, that it is equally surprising that the grey matter in the brain brings about conscious experiences, and the brain is probably the most plausible candidate to do it.

The differences in the force of the intuition are obviously biased by the fact that we have a brain and that we have phenomenally conscious experiences. When one reflects on it, one can find equally mysterious how experiences are produced by our brain (which is precisely the hard problem of consciousness).

Additionally, one could argue that we wouldn't *see* any phenomenally conscious experience in the Chinese nation and we could explain the behavior of the artificial body without appealing to phenomenally conscious experiences. But there is nothing new here. The very same thing happens in the case of the brain: one cannot *see* any of my phenomenally conscious experiences in my brain and, knowing enough neuropsychology, one can explain my behavior without appealing to the phenomenal character of my experiences.

For my part, I favor the functionalist intuition. This intuition has been nicely motivated by Chalmers (1995). Chalmers' thought experiment proposes us to imagine a silicon chip that is functionally identical to a neuron and a subject whose neurons are replaced by these chips. We will produce a functional duplicate for every specific kind of neuron. If the brain has m neurons that can be divided on the basis of their function (number of connections, firing frequency margins, etc.) into n kinds, we will have to develop n different kinds of silicon chips, one for each functionally distinguishable neuron type.

A subject called Joe volunteers for the experiment. Just before starting the experiment, Joe can see a big flag at the end of the room in bright red and yellow colors. Due to stress, he feels a horrible headache before the experiment starts. The experiment goes as follows: Joe's neurons are replaced one by one by the corresponding silicon chips. At the end of the process, Joe's brain will be a silicon brain. At that time, either he is phenomenally conscious or he isn't. If he is phenomenally

realizing one functional state in such a theory. In my presentation of the example, I am considering a theory at the level of neural communication because this is the most basic level that cognitive neuroscience usually considers.

conscious, then there is no problem for the functionalist. If he isn't, then either at a precise neuron replacement Joe loses phenomenal consciousness (the pain and the visual experience suddenly disappears) or his phenomenally conscious experience gradually fades away with each replacement.

The question at this point is not whether we can know whether Joe has a conscious experience or not. Being a functional duplicate of himself before the operation, he will reply exactly the same as he would have replied if there were no replacement and we have no access to his experience beyond his reports. The knowability from our perspective is irrelevant; what matters is whether Joe does or doesn't feel something. The question is not epistemic (see the Harder Problem below) but metaphysical.

In the case of a suddenly disappearing sensation, a replacement in a single neuron would make Joe's horrible headache vanish. Moreover, the mental thought experiment can be replicated at a finer-grained level, at the level of molecules instead of neurons for example, to a point where a single molecule replacement makes the pain sensation suddenly disappear. The idea that a single molecule replacement makes the sensation suddenly disappear seems to be implausible.

Alternatively, one could maintain that Joe's conscious experiences fades away with each replacement. This option is not better, for it requires Joe to be systematically wrong about his experience while everything is functionally right. In this case the phenomenal character of Joe's experience fades away with every replacement, but he always reports having a horrible headache. Joe would be systematically wrong about his experience, but *ex-hypothesi* he is functionally perfectly right and ideal –as right and ideal as he was before the replacement started.

> Joe is systematically wrong about everything that he is experiencing. He certainly says that he is having bright red and yellow experiences, but he is merely experiencing tepid pink. [...] Worse, on a functional construal of judgment, Joe will even judge that he has all these complex experiences that he in fact lacks.
>
> [...] There is a significant implausibility here. This is a being whose rational processes are functioning and who is in fact conscious, but who is completely wrong about his own conscious experiences. [...] In every case with which we are familiar, conscious beings are generally capable of forming accurate judgments about their experience, in the absence of distraction and irrationality. For a sentient, rational being that is suffering from no functional pathology to be so systematically out of touch with its experiences would imply a strong dissociation between consciousness and cognition. We have little reason to believe that consciousness is such an ill-behaved phenomenon, and good reason to believe otherwise. Chalmers (1995)

We have seen that the intuition that the *Chinese Nation* would lack consciousness seems to be supported by the mystery of consciousness itself and not to be related to anything that is particular to the functionalist approach. As I have noted, it is equally surprising that our brain gives rise to phenomenally conscious experiences.

Furthermore, Chalmers' argument is surely not conclusive, and not strong enough to convince anti-functionalists, but it makes it very plausible that having a phenomenally conscious experience is a matter of being in a state that satisfies a certain functional role, which is independent of its realization.

The Harder Problem of Consciousness

In chapter 1 I presented the hard problem of consciousness. The hard problem is the problem of explaining why and how a physical system gives rise to phenomenally conscious experiences. The harder problem of consciousness is an argument introduced by Block (2002a) to show an epistemic problem for the materialist position.

In the original paper, Block mentions an episode of the TV show 'Star Trek: The Next Generation' where there is a trial to decide whether it would be legal to turn off and take apart Commander Data (an android that looks and behaves like a human) by someone who doesn't know whether the parts can be put together again.[8] In the end, the decision comes down to whether or not Commander Data is phenomenally conscious. He behaves exactly as a human, at a superficial level he is a functional duplicate of a human. We have no good reason, apart from chauvinism, to believe that phenomenal consciousness is not multiply realizable. In fact, following Chalmers (1995), I have given reasons to believe that this is the most plausible option.

The problem for materialism is the following: we have reasons to believe that Commander Data is phenomenally conscious. For instance, he claims that he feels pain when shot. But these reasons are easily defeated; the fact that Data reports that he is in pain when shot seems not to be enough for ascribing to him the corresponding phenomenally conscious experience. If this is right, then we have no way to decide whether Data is phenomenally conscious or he isn't, given that he does not relevantly share our microphysical nature.

For beings that are physically similar to us like animals, one can appeal to analogy to establish whether they are phenomenally conscious or not. Such a possibility is not available in the case of Data. There are a bunch of properties that Data and I share, but we have no empirical way to decide whether these properties suffice for having phenomenally conscious experiences.

I think that the harder problem of consciousness is the problem of individuating the properties that are essential for having a phenomenally conscious experience. Let me elaborate:

Imagine that N is the neural correlate of a phenomenal property Q. N is the brain activity that is minimally sufficient for having an experience with phenomenal character Q. Whenever I instantiate N I undergo a phenomenally conscious experience with phenomenal character Q. If P is the microphysical implementation of N at the level of fundamental physics, for example a collection of strings if strings are the fundamental particles of Physics, then I instantiate P. If F_i is the implementation of N at the level of atoms then I also instantiate F_i. P implements also F_i, atoms are made out of strings. If F_i could have had a different implementation at the level of the fundamental particles of Physics, namely, if atoms do not have a unique implementation at the level of fundamental physics, then the property of having P and the

8 See footnote 17 in Block's article for a synopsis of the episode.

property of having F_i are different properties. I also instantiate F_j, the implementation of N described at the level of chemical activity among the neurons, via synapsis, that conform N (F_i implements F_j) and I also instantiate F_k, which is the implementation of N at the level of computational communication between the neurons that conform N.

N has different properties qua network of strings, qua network of atoms, qua neuro-chemical network, qua computational system, etc. There is a hierarchy of properties depending on the level of abstraction. The satisfaction of a lower level of abstraction guarantees the satisfaction of a higher-level of abstraction, because lower levels of abstraction are realizers of (implement) higher-levels of abstraction. For instance, if a subject satisfies F_i (the atomic level) then she will satisfy F_j (the neuro-chemical level) and F_k (the computational neural level). On the other hand, the satisfaction of a higher-level of abstraction does not entail the satisfaction of a lower level, if there can be multiple realizability. For instance, someone could satisfy a certain computational neural level (she satifies F_k) without thereby satisfying the neuro-chemical level (she does not satisfy F_j) because she has a silicon brain. The way neurons exchange information is relevant for the instantiation of F_k but it is not so for F_j. Neurons perform such an exchange via synapsis, silicon chips via electrical signals instead. We don't want to claim that every property instantiated in virtue of having N is necessary for phenomenal consciousness.

How can we select the properties that are necessary for phenomenal consciousness? Commander Data's brain and mine share a bunch of properties; however, there is another bunch of properties they don't share. The harder problem puts forward the epistemic problem materialists have to face: once the description at a certain superficial level is satisfied by a subject (for instance, if she is behaviourally indistinguishable from me), there is no way of empirically testing whether such a subject has or not phenomenally conscious experiences if the physical implementation is relevantly different from paradigmatic cases. Commander Data satisfies this superficial level and he behaves in circumstances C as a human would behave in C. However, he is relevantly different from us and we cannot be sure whether he has phenomenally conscious experiences or not. There is no empirical way to decide among the different levels. If we find the neural correlate of a certain phenomenal property, at which level of description should we look for this phenomenal property? (i.e., which properties of this neural correlate are essential to phenomenal consciousness?): the molecular ones? The neural ones? The neuro-chemical ones? There seems to be no empirical way to answer to these questions.

One could claim that we can only be sure about the microphysical level. We know that our molecules give rise to consciousness but we cannot possibly know about a different implementation. The problem of this approach is that phenomenal consciousness would depend on the fundamental particles of the physics that may be the same for all kind of macrophysical things. The fundamental particles that constitute all the things in the actual world would have certain properties –proto-phenomenal properties. Once these fundamental particles are organized in a certain manner (as they are organized in N for instance) they will give rise to phenomenal properties (an experience as of red in the case of N). I find this form of proto-panpsychism unappealing.[9]

9 This form of proto-panpsychism has been defended by Chalmers (1996).

My inclination is toward the highest possible level, exactly the opposite. For that purpose we need to find a functional characterization of phenomenal consciousness; a functional characterization that gives us a criterion for ruling out certain properties of the neural correlate that are not necessary for consciousness. I maintain that if Commander Data satisfies this function then he has phenomenally conscious experiences. This inclination is supported by the intuition presented by Chalmers that if the functional description is guaranteed then phenomenal consciousness is preserved. The harder problem shows that there is no way to empirically test this hypothesis.

Furthermore, we have to inquire, relying on *a priori* reasoning, empirical evidence and phenomenological observation, the conditions under which phenomenal conscious experience exists. We are not going to find the neural correlate until we agree on the level that gives rise to consciousness. For instance, a theory that maintains that the cognitive access underlying reportability is necessary for phenomenal consciousness[10] will postulate a different neural correlate than a theory that suggests the opposite. The former will maintain that the neural activity in the frontal lobe is essential to a phenomenally conscious mental state whereas the latter could deny it.

In chapters 4 and 5 I will discuss some considerations for what I consider a proper functional account of phenomenal consciousness.

Functionalism and physicalism

In the previous sections, I have deliberately chosen to use the term 'materialism' instead of 'physicalism' despite the fact that they are commonly used as synonyms. In chapter 1, I introduced materialism as the thesis that everything in the actual world is itself physical or metaphysically supervenes on physical facts. Several positions are covered under the umbrella of this description, some of them rivals. Block (ming) has suggested that only some of them deserve the name 'physicalism'. In this subsection, following Block, I will present these positions and the tension between them. I will make it clear that my interest in this chapter is to defend materialism, as presented in the previous chapter, with regard to certain objections.

There is a distinction between Ontology on one side and Metaphysics on the other. Block, following the use that Quine (1948) makes of the terms, defines Ontology as the study of the kind of things that exist and contrasts it with Metaphysics as the study of the ultimate nature of things. For instance, metaphysical theories of numbers provide different accounts of what numbers are: for platonists they are abstract objects in a third-realm; for constructivists, numbers are mind-dependent objects, they exist only in so far as they can be constructed; for structuralists they are elements within a structure, etc. The ontology of a particular entity refers to what it is made of, what its existence really entails. In the case of platonism, the existence of numbers commits us to a third-realm, in the case of constructivism to the existence of minds, etc.

With this distinction in hand, we can look back at the problems we are facing. In first place, there is an ontological problem: does the acceptance of the reality of phenomenally conscious experience commit us to something that does not metaphysically supervene on physical properties? The ontological physicalist replies that it doesn't,

10 This supposition, common to many scientific studies, is criticized in 5.2.2.

the ontological dualist replies that it does. They disagree about the kind of things that (metaphysically) exist.

Functionalism is a metaphysical thesis. Functionalism maintains that the ultimate nature of the properties in virtue of which a mental state is a phenomenally conscious mental state is a certain functional role; phenomenal properties are functional properties. On the other hand, what Block calls metaphysical physicalism is the thesis that phenomenal properties are themselves physical properties. They do not merely metaphysically supervene on physical properties but they are identical to physical properties. According to metaphysical physicalism, phenomenal properties are themselves physical; they are not just properties that metaphysically supervene on physical properties. Metaphysical physicalism does not admit multiple realizability. Functionalism and metaphysical physicalism are rival theses.

Surely, one can be a metaphysical physicalist about certain properties and a metaphysical functionalist about others. One can be a metaphysical functionalist about aesthetics or economics and a metaphysical physicalist about the mental.

Functionalism is compatible with both ontological physicalism and dualism. Functionalism is a metaphysical thesis; what is further required is an ontological thesis. What I have called materialism is safe insofar as ontological physicalism is true; what materialism requires is ontological physicalism; it is therefore compatible with both metaphysical physicalism, which entails ontological physicalism,[11] and metaphysical functionalism combined with ontological physicalism.

Block complains that if metaphysical functionalism is true about the mental, then the project of reducing the mental to the physical fails in an important sense:

> My point is that the reductive physicalist needs both. Suppose the ontological concerns of functional reduction are satisfied. All realizers—even all possible realizers—of mental roles are physical. Still, if there is no physical property in common to your realization of phenomenal quality Q and mine that can explain or constitute the similarity, if the only common property that can explain the similarity is functional, then in an important sense physicalism about properties is false. It is that fact that is left out by the functional reduction point of view. (ibid, p.36)

Be that as it may, I am not interested here in this form of reductionism or in issues of physicalism. What I will be addressing in this chapter is how to defend ontological physicalism from anti-materialist arguments, and in the rest of the dissertation I will try to offer a theory that is compatible with it. Metaphysical functionalism is compatible with ontological physicalism. If phenomenal properties are functional properties and the realizers in the actual world are physical then a minimal physical duplicate of the actual world is a duplicate simpliciter of the actual world and, therefore, materialism would be true in our world. Functionalism is therefore an acceptable position for my purposes which is to clarify whether and how the physical can give rise to phenomenal consciousness.

11 In the same way metaphysical dualism entails ontological dualism.

2.1.3 *The Modal Argument Raised: Zombies.*

Functionalists who are materialists maintain that phenomenally conscious experiences metaphysically supervene on the physical. Kodos' pain is not a problem for functionalism; pain is a functional property that is multiply realizable. Unfortunately for the materialist, avoiding identification between phenomenal properties and physical properties is not enough for blocking the modal argument. Even if one does not want to hold an identity thesis but a supervenience thesis, the problem reemerges. If the phenomenal property Q metaphysically supervenes on a set of microphysical properties P (remember that if P is fixed, so are the higher functional levels), then it is metaphysically necessary that the conditional $P \rightarrow Q$ is true. If P and Q are rigid designators, then if $P\&\neg Q$ is conceivable we can build up a new modal argument.

This section is about this argument. I will start by presenting the argument in a simplistic way. We will see that the argument is based on a link between conceivability and metaphysical possibility. In order to make such a link plausible, the notion of conceivability has to be refined. Chalmers (2002) presents a taxonomy with three dimensions along which conceivability can vary, I will introduce such a taxonomy and I will then formally present the argument.

The conceivability of P without Q has been famously defended by Chalmers (1996). In order to pump the intuition, Chalmers introduced a new character in the philosophy of mind: philosophical zombies. A philosophical zombie is not a reanimated corpse, nor a human being who is controlled by someone else through the use of magic, nor the victim of a government's experiment causing a weird pandemic. A philosophical zombie looks and behaves like you and me. A zombie is a hypothetical creature that is a microphysical duplicate of a phenomenally conscious being such that it lacks any phenomenally conscious experiences.

Chalmy is a microphysical duplicate of David Chalmers in exactly this moment, t_o. In t_o they both satisfy exactly the same microphysical description P, and with it every functional description F.[12] If, for instance, some time later they are presented with different stimuli, then they will differ microphysically. They will, however, behave exactly in the same way in the same circumstances, at least in the beginning.[13] They both will go for a walk with friends and have a beer commenting on how nice is to enjoy the delicious beer and a good philosophical discussion. What is relevant is the big difference between Chalmers and Chalmy: the latter has no phenomenally conscious experiences at all, there is nothing it is like for Chalmy to taste the beer. Chalmers argues that if Chalmy is conceivable then materialism is false, for the same reasons given by Kripke. It can be the case that P is true but Q is false (Chalmers 2002). Formally:

> Let P be a complete microphysical description of the world;
>
> Let Q be a phenomenal truth, for instance the one expressed by the sentence: 'Chalmers has a headache'.

12 I am considering notions of function that are structure dependent. In theories such as etiological ones, the function of a system depends not only on its structure, but also on its historical properties. In that case we can assume that the evolutionary history of Chalmers and Chalmy is identical. I will deal with etiological theories in 4.4.

13 As they learn different things their behavior under the same circumstances will differ.

Materialism, as we have seen in the first chapter, holds that phenomenal properties metaphysically supervene on physical properties. Materialism is therefore committed to the idea that it is metaphysically necessary that $P \rightarrow Q$:

(Anti-materialism)

(1) If materialism is true, then $P \rightarrow Q$ is metaphysically necessary.

(2) $P\&\neg Q$ is conceivable (we can conceive of the falsity of the conditional $P \rightarrow Q$).

(3) If $P\&\neg Q$ is conceivable then $P\&\neg Q$ is metaphysically possible.

(4) It is metaphysically possible that $P\&\neg Q$ (from 2 and 3).

(5) It is not metaphysically necessary that $P \rightarrow Q$ (from 4).

$\therefore$ Materialism is false (From 1 to 5)

In order to support an entailment thesis between conceivability and metaphysical possibility, (3), the idea of conceivability has to be refined. No one would accept it as presented above. Some things are conceivable due, for instance, to the limitations of our cognitive system, without thereby being metaphysically possible. Chalmers (2002) presents a taxonomy with three dimensions along which conceivability can vary:

PRIMA FACIE VS. IDEAL CONCEIVABILITY

Someone finds a statement *prima facie* conceivable if, and only if, it is conceivable after certain consideration. An example of prima facie conceivability could be a very complex mathematical truth. I will call it Mat. Despite being provable, most people would find it prima facie conceivable that Mat is false. Furthermore, a situation described by a very long sentence could be prima facie inconceivable, for we lack the memory or other cognitive resources necessary to parse it (Chalmers, 2002, p. 147).

On the other hand, it is not ideally conceivable that Mat is false, and the very long sentence is ideally conceivable. Ideal conceivability abstracts from cognitive limitations by introducing the notion of an ideal conceiver, who has no cognitive limitation. A statement is ideally conceivable if, and only if, the ideal conceiver can conceive it. Mathematical truths and long sentences are not a problem for our ideal conceiver. She has unlimited memory and she can get a proof of any mathematical truth.

NEGATIVE VS. POSITIVE CONCEIVABILITY

A statement S is negatively conceivable if, and only if, it cannot be ruled out a priori. The statement is prima facie negatively conceivable for a subject S when S, after consideration, cannot rule out S on a priori grounds. And we can say that S is ideally negatively conceivable when it is not a priori that S is not the case. (Chalmers, 2002, p. 147)

The idea of positive conceivability requires some form of imagination:[14]

14 The notion of positive conceivability was introduced by Yablo (1993).

> Positive notions of conceivability require that one can form some sort of positive conception of a situation in which S is the case. One can place the varieties of positive conceivability under the broad rubric of imagination: to positively conceive of a situation is to in some sense imagine a specific configuration of objects and properties ... Overall, we can say that S is positively conceivable when one can imagine that S: that is, when one can imagine a situation that verifies S. (Chalmers, 2002, pp. 147-148)

This form of imagination is supposed to be a special faculty of modal imagination that transcends imagery so that one may modally imagine "pairs of situations that are perceptually indistinguishable" and situations that are "unperceivable in principle." (Chalmers, 2002, p. 149)

PRIMARY VS. SECONDARY CONCEIVABILITY

The distinction between primary and secondary conceivability comes from Kripke's discussion of a posteriori necessities. Primary conceivability intends to capture the sense in which a posteriori necessities like 'water is H_2O' are conceivable. Primary conceivability is an epistemic possibility for a competent, but sufficiently clueless speaker. A statement S is primarily conceivable when it is conceivable that S is actually the case. Primary conceivability is an epistemic notion; it is "is grounded in the idea that for all we know a priori, there are many ways the world might be."

Let me refer to any substance that satisfies the contingent properties that help fixing the reference of the term 'water' as *watery stuff.* Watery stuff will be any substance that satifies a description like the following: 'the liquid that comes from the tab, filling lakes and rivers, etc'. According to Kripke, we call the watery stuff in our world 'water' and a posteriori we discover that water is H_2O. Consider an English speaker of the 15th century or a current speaker lacking any chemical knowledge; for all they know, water might be a different substance, say XYZ; in other words, the contingent properties that fix the reference of a term like 'water' might be satisfied by another substance. There is a sense in which it is conceivable that water is not H_2O: it is primary conceivable that water is not H_2O.

On the other hand, secondary conceivability relates to ways the world might have been but it is not. This is the sense in which, if Kripke is right, something cannot contain water and fail to contain H_2O, for water is H_2O. It is not secondarily conceivable that water is not H_2O. In this sense a situation in which we conceive of water being XYZ should be better described as a situation in which water is H_2O, but the watery stuff filling lakes is XYZ. This notion of conceivability is not very interesting for anti-materialist's purposes for materialists will deny that zombies are secondarily conceivable.

Parallel to the distinction between primary and secondary conceivability there is a distinction between primary and secondary possibility. A statement S is primarily possible, if and only if it is true in every possible world considered as actual. S is secondarily possible, if it is true in every possible world considered as counterfactual. If the actual world were such that the watery stuff (the liquid coming from the tab, filling lakes and rivers, etc) were XYZ, then 'water' would refer to XYZ. It is primarily possible that water is XYZ.

The actual world is such that the watery stuff is H_2O, then if Kripke is right, there is no possible world considered as counterfactual in

which water is XYZ. It is not secondarily possible that water is XYZ. Secondary possibility is metaphysical possibility.

It seems plausible to maintain that primary conceivability entails primary possibility, but cases like water being XYZ are primarily conceivable but not secondarily possible. Primary conceivability is an imperfect guide to metaphysical possibility. It is not always the case that if a situation is primarily conceivable it is metaphysically possible, as the previous example shows. According to Chalmers, this is so because in the case of water primary possibility does not entail secondary possibility. However, in the case of phenomenal consciousness primary possibility entails secondary possibility because, as Kripke argues, the properties that help fixing the reference in the case of terms that refer to phenomenal properties are essential to the phenomenal properties.

With this tool in hand, Chalmers (2002) presents a refined version of (anti-materialist), that builds on the ideal primary conceivability of zombies.[15]

(Anti-materialist Chalmers)

(1) If materialism is true, then $P \rightarrow Q$ is necessary.

(2) $P\&\neg Q$ is ideally primarily (positively/negatively) conceivable.

(3) If $P\&\neg Q$ is ideally primarily (positively/negatively) conceivable then $P\&\neg Q$ is primarily possible.

(4) If $P\&\neg Q$ is primarily possible, then $P\&\neg Q$ is secondarily possible.

(5) It is not necessary that $P \rightarrow Q$ (from 2, 3 and 4)

---

(6) Materialism is false (From 1 to 5)

Materialists have several possibilities to block (anti-materialist Chalmers):

The first one is to deny (2): zombies are not primarily conceivable. They may seem conceivable but they are not; on reflection, there is no further problem for explaining phenomenal consciousness beyond explaining the various cognitive, behavioral, and environmental functions. There is no 'hard problem' of consciousness beyond the 'easy problem'. This is what Chalmers (2003a) calls type-A materialism. Type-A materialists either embrace eliminativism and deny that there is such a thing as phenomenal consciousness or maintain that there is an *a priori* connection between phenomenal concepts and physical concepts. There is an *a priori* entailment between physical truths and phenomenological truths that prevents the conceivability of zombies: zombies are not ideally negatively conceivable. Examples of type-A materialism are Harman (1990); Dennett (1991); Lewis (1994); Rey (1995); Churchland (1996).

Type-B materialists deny the link between conceivability and metaphysical possibility. Type-B materialists accept that zombies are conceivable in a relevant sense (they accept that there is no logical contradiction

15 I want to leave it open here, as Chalmers does, whether positive or negative conceivability is required for the premises of the argument. I will come back to this point in 2.1.4. Suffice it to say that ideal primary positive conceivability entails ideal primary negative conceivability and that the former is a better guide for metaphysical possibility.

in the idea of a zombie) but deny that such a conceivability is always a good guide to metaphysical possibility. They accept the existence of an epistemic gap between physical truths and phenomenal truths that guarantees the negative conceivability of zombies, but they deny the ontological gap. Type-C materialism is like type-B materialism, but it maintains that such an epistemic gap is only temporary, that it will be closed. [16]

In the next subsection I will present what I consider to be a convincing reply to the former argument.

2.1.4 *A Materialist Reply to the Modal argument*

Balog (1999) has argued that accepting the same premises that support the argument (the conceivability of zombies and the link between conceivability and possibility) we can build up a valid argument with an unacceptable conclusion. This constitutes a good reason for rejecting the link between conceivability and possibility.

Balog considers a zombie world, a microphysical duplicate of our world but lacking phenomenal consciousness. In the zombie world materialism is true. To say that materialism is true in a world w is to say that everything there is in w metaphysically supervenes on the physical things. If materialism is true in a world w, then a minimal

16 Materialists could endorse a form of proto-panpsychism to block premise 4. If Kripke is right, Q poses no problem to (4), because anything that is counterfactually felt like Q cannot fail to be Q. For those familiar with the bidimensional framework, this is equivalent to the claim that primary and secondary intentions of Q coincide; primary and secondary possibility coincide in the case of phenomenal consciousness. But we might not be able to say the same about P. Chalmers has convincingly argued that materialists can resist (4) by endorsing a form of proto-panpsychism (Feigl (1958); Chalmers (1996); Stoljar (2005)) in which "consciousness is constituted by the intrinsic properties of fundamental physical entities." (Chalmers (2002, p. 265)).

> [A] world can verify P without satisfying P, it may be that P&~Q is 1[primarily]-possible but not 2[secondarily]-possible. However, this requires that P and Q be related in a certain specific way. In particular, it requires that some worlds that verify P also verify ~Q, while no worlds that satisfy P also satisfy ~Q. This requires in turn that some worlds that have the same structural profile as the actual world verify ~Q, while no worlds that have the same structural and intrinsic profiles as the actual world satisfy ~Q. We can assume for the moment that the primary and secondary intensions of Q coincide. Then we can put all this by saying that the falsity of [(4)] requires that the structural profile of physics in the actual world does not necessitate Q, but that the combined structural and intrinsic profiles of physics the actual world do necessitate Q.
>
> This idea — that the structural properties of physics in the actual world do not necessitate the existence and/or nature of consciousness, but that the intrinsic properties of physics combined with the structural properties do — corresponds to a familiar view in the metaphysics of consciousness. This is the view that I have elsewhere called Russellian monism (or type-F monism, or panprotopsychism). On this view, consciousness is closely tied to the intrinsic properties that serve as the categorical bases of microphysical dispositions. Russell and others held that the nature of these properties is not revealed to us by perception (which reveals only their effects) or by science (which reveals only their relations). But it is coherent to suppose that these properties have a special nature that is tied to consciousness. They might themselves be phenomenal properties, or they might be protophenomenal properties: properties that collectively constitute phenomenal properties when organized in the appropriate way. Chalmers (2009)

I do not find proto-panpsychism to be an attractive alternative for materialism. It is not clear to me that a theory that postulates such proto-phenomenal properties can be considered a 'suitable improvement' of our current physics. Be that as it may, the purpose of this work is not to argue against this position. In this chapter I try to offer an alternative that is compatible with materialism and is not committed with proto-panpsychism.

microphysical duplicate of w is a duplicate of w simpliciter. If a zombie world is metaphysically possible then materialism is false of the actual world, for the zombie world is a minimal microphysical duplicate of the actual world and is not a duplicate of the actual world simpliciter.

The claim that materialism is true in our world is compatible with there being a microphysical duplicate of our world w' but containing angels (non-physical entities), because w' would not count as a minimal physical duplicate. Materialism would be false in w'. *Ex-hypothesi*, in the zombie world, materialism is true.

Chalmy, the zombie twin of David Chalmers, is a microphysical and functional duplicate of David Chalmers. In (Zombie Anti-materialism), Q* expresses a true thought in the zombie world that corresponds to the one expressed by Q in the actual world. For instance, if Q expresses Chalmers' thought that he is having a SENSATION OF PAIN in the actual world, Q* expresses the thought that Chalmy is having a SENSATION OF PAIN* in the zombie world. These two thoughts are equivalent. By that I mean that whenever Chalmers tokens SENSATION OF PAIN, Chalmy will token the equivalent SENSATION OF PAIN*. The concepts SENSATION OF PAIN and SENSATION OF PAIN* have a completely parallel role in their beliefs, desires, cognition etc. Chalmers and Chalmy are functionally indistinguishable. When Chalmers (Chalmy) thinks that he is in pain (pain*), he talks to his friend. His friend recommends that he visits the doctor (zombiedoctor). The doctor (zombiedoctor) asks Chalmers (Chalmy) where is he feeling (feeling*) pain (pain*) and thanks to Chalmers's (Chalmy's) report the doctor (zombiedoctor) discovers a small problem that despite being easily operable could have had horrible consequences.[17] This is something that the conceivability of functional duplicates lacking phenomenal consciousness requires, something that I will concede to the advocate of the modal argument.

Chalmy reasons as follows:[18]

(Zombie Anti-materialism)

(1*) If materialism is true, then $P \rightarrow Q*$ is metaphysically necessary.

(2*) $P\&\neg Q*$ is conceivable (zombies can conceive the falsity of the conditional $P \rightarrow Q*$).

(3*) If $P\&\neg Q*$ is conceivable then $P\& \sim Q*$ is metaphysically possible.

(4*) It is metaphysically possible that $P\&\neg Q*$ (from 2* and 3*).

(5*) It is not metaphysically necessary that $P \rightarrow Q*$ (from 4*).

(6*) Materialism is false (from 1 to 5).

17 Someone could complain that Chalmy's thoughts are meaningless, for the concepts involved do not refer. If thoughts play any role in cognition and behaviour then Chalmy's thoughts cannot be meaningless, for Chalmers and Chalmy are functionally identical as the previous story illustrates.

18 The argument is adapted from Balog (mingb).

(Zombie Anti-materialism), as the original anti-materialist argument, is a valid argument. If (anti-materialism) is sound, so is (Zombie Anti-materialism). The conclusion in this case is, *ex-hypothesi*, false: materialism is true in the zombie world. It is metaphysically necessary that every world where P obtains is a world where Q* obtains. Consequently, one of the premises has to be false.

We can reject (2*) or (3*).[19] When Chalmers, in the actual world, considers that there could be a microphysical duplicate of him that doesn't feel any pain, Chalmy considers, in zombie-world, that there could be a microphysical duplicate of him that doesn't feel any pain*. If there is a conceptual disconnection between P and Q*, then P&¬Q$*$ would be negatively ideally primarily conceivable, because there wouldn't be a logical contradiction in the idea of a zombie that lacks Q*. In section 2.3 I will develop this idea and show that if P and Q are not conceptually connected nor are P and Q*. This will support (2*).

If (3*) is false, then there is no link between conceivability and possibility. However, if this is a case of a failure in the connection between conceivability and possibility, then (Anti-materialist Chalmers) is unsupported. The dualist is left with no reason to believe that conceivability entails metaphysical possibility.[20]

Unfortunately for the materialist, there is one possibility left to the dualist. She can claim that whereas we can positively conceive P&¬Q, zombies cannot positively conceive P&¬Q$*$. Dualists can suggest that the notion of conceivability involved in the argument (positive vs. negative) should not be left open as in (Anti-materialist Chalmers). Metaphysical possibility requires positive conceivability. The kind of conceivability involved in (Anti-materialist Chalmers) and in (Zombie anti-materialism) is different: both P&¬Q and P&¬Q$*$ are negatively conceivable, but only P&¬Q is positively conceivable.

Neither the truth of ¬Q$*$ nor the truth of ¬Q can be ruled a priori by a microphysical description of the world, P. The conceptual disconnection between Q (Q*) and P guarantees the negative conceivability where only a priori reasoning is required.[21]This will be enough for leaving $3^{negative}$ unsupported.

$3^{negative}$ If P&¬Q is primarily negatively ideally conceivable, then P&¬Q is metaphysically possible.

However, P&¬Q is positively conceivable, whereas P&¬Q$*$ is not. $3^{positive}$ is left untouched.

$3^{positive}$ If P&¬Q is primarily positively ideally conceivable, then P&¬Q is metaphysically possible.

The advocate of dualism could maintain that primary positive ideal conceivability and not primary negative ideal conceivability entails metaphysical possibility and that, whereas P&¬Q is positively conceivable, P&¬Q$*$ is only negatively conceivable.[22] The conceivability

19 Of course, one could also deny at this point the conceivability of zombies (type-A materialism), but, as I said, I am going to concede this premise to my opponent.

20 There is an interesting reply presented by Chalmers (2009). The corresponding concepts of phenomenal concepts (phenomenal* concepts) in the zombie world do not refer, he argues. Q* is either false or meaningless and the conceivability of P&¬Q$*$ is therefore not incompatible with materialism. However, it is hard to see how, in such a case, Chalmy can be functionally identical to Chalmers. In section 2.3, where I present the phenomenal concept strategy, I will defend it against Chalmers' objection and maintain that under most current theories of mental content, phenomenal* concepts do refer.

21 Materialists have to explain this exceptional disconnection. I will address this issue in 2.2.

22 Note that positive conceivability entails negative conceivability.

argument against materialism remains untouched unless the materialist can argue that P&¬Q∗ is positively conceivable.

Balog (mingb) acknowledges this, but instead of arguing in favor of the positive conceivability of P&¬Q∗, she provides a new argument. This argument shows that in order to block the anti-materialist argument (even if one accepts that P&¬Q is primarily positively ideally conceivable) it is enough to maintain that it is negatively conceivable that purely physical beings have phenomenally conscious experiences. The argument is the following:

(A priori anti-dualism)

(1) It is a priori that P&¬Q is primarily positively ideally conceivable

(2) It is a priori that if P&¬Q is primarily positively ideally conceivable, then P&¬Q is metaphysically possible

(3) It is a priori that if Q is true and P&¬Q is metaphysically possible, then materialism is false.

---

(4) It is a priori that if Q is true, then materialism is false.

The positive conceivability of P&¬Q has to be granted by the dualist, because it is required by the version of the anti-materialist argument that remains immune to the objection presented by (zombie anti-materialism). This version is based on the positive conceivability of zombies. The a priority required by the first premise is justified by the conceptual independence of phenomenal and physical descriptions (if they were connected then zombies wouldn't be conceivable).

The second premise is just an extension of the entailment between positive conceivability and metaphysical possibility that the dualist suggests. If the entailment is true, it is true a priori.

As regards premise (3), it cannot be a priori that if P&¬Q is metaphysically possible, then materialism is false. If there were no phenomenal consciousness, then materialists would have nothing to fear. However, we are considering realist positions with regard to phenomenal consciousness (eliminativism is not an opponent at this point): there is phenomenal consciousness in the actual world, and therefore only those possible worlds where there is phenomenal consciousness are relevant. Materialism is uncontroversially true of all worlds that are close enough to the actual world in which there is no phenomenal consciousness. The fact that Q is the case in certain worlds cannot be known a priori. For that reason, in order to get an a priori statement, the truth of Q is added in (3).

From these three premises we can derive the conclusion (4) that it is a priori that if Q is true then materialism is false.[23] If the argument is valid, the only thing that the materialist has to do to reject that metaphysical possibility is entailed by positive ideal primary conceivability is to show that the conclusion of (A priori anti-dualism) is false. How can the materialist show that (4) is false? Given the link between a priori reasoning and negative conceivability, the statement 'it is a priori that if Q is true, then materialism is false' is equivalent to this one: 'It is

23 Brown (2010) has proposed a similar line of argument.

not negatively conceivable that Q is the case and materialism is true'. The materialist just needs to support the thesis that it is negatively conceivable that materialism is true and Q is the case.

Consider an anti-zombie world: a world that is a microphysical duplicate of ours, lacking any non-physical property. The anti-zombie world is inhabited by anti-zombies. Anti-zombies have phenomenally conscious experiences. Materialists actually maintain that the actual world is an anti-zombie world, but the argument doesn't require that thesis to be true, it doesn't even require the metaphysical possibility of an anti-zombie world, only that it is negatively conceivable.

Chalmo, is an anti-zombie. He is a micro-physical duplicate of David Chalmers, has conscious experiences and is purely physical.[24] Materialists don't have to argue that anti-zombies are positively conceivable. If Chalmo is negatively conceivable then the materialist can deny the link between conceivability and metaphysical possibility.

Negative conceivability requires that one cannot rule out anti-zombies a priori, there not being contradiction in the idea of anti-zombies. The conceivability of zombies is based on the conceptual independence of phenomenal and physical concepts. The same reason that allows the conceivability of zombies guarantees the negative conceivability of anti-zombies. There is no logical contradiction in the idea that phenomenal properties are identical with certain physical or functional properties and therefore anti-zombies are negatively conceivable.

The remaining task for the materialist in order to reply to the modal argument is to explain the conceivability of zombies and anti-zombies. I will offer such an explanation in section 2.3 where I will present the Phenomenal Concept Strategy. In a nutshell, the idea is that if there is no conceptual connection between P and Q then, even if Q were something physical, there would not be a logical contradiction in $P\&\neg Q$ and therefore $P\&\neg Q$ would be negatively conceivable.

Let me first present a second and related argument against materialism: the knowledge argument and the related explanatory gap.

2.2 THE EXPLANATORY GAP

Jackson (1982) has presented the argument that probably best illustrates the hard problem. The argument is known as the *Knowledge argument*.

Jackson invites us to imagine Mary, a scientist who was captured at birth by a mad scientist and enclosed in a black and white room. She has contact with the external world only through black and white books and black and white television. In a modern version of the experiment we can imagine that the kidnapper puts special contact lenses in Mary's eyes that filter the light in such a way that she can only see in black, white and different shades of grey. Mary's visual experience of the world is similar to the one we had while watching *Citizen Kane*. Mary studies color and color vision. She becomes an expert in color vision to the point that she comes to know all the physical facts pertinent to it; how objects reflect the light, how light excites the eye cells, how the signal is transmitted through the optic nerve, all the processes corresponding to color vision, and whatever the latest research can tell us about color vision.

24 Purely microphysical duplicates that enjoy phenomenal consciousness have been discussed in the literature; Balog calls them illuminati, Brown (2010) calls them shombies, etc.

When Mary is released and her contact lenses are removed, she sees a rose and she exclaims: "So, that is what it is like to experience red!". Intuitively, it seems that she learns something by seeing red for the first time, namely what it is like to see red, the way undergoing the experience *feels*. However, ex-hypothesi, Mary already knew all the physical facts relevant for color vision. The thought experiment reveals a gap in the explanation of phenomenal properties in terms of physical properties no matter how deep our physical knowledge might be. No matter how deeply we reflect, we cannot acquire phenomenal knowledge (that I am having a phenomenally conscious experience as of red, for instance) from knowledge of physical facts.

The denial of the link between conceivability and possibility discussed in the previous section does not solve all of the materialist's problems. The failure of the entailment of phenomenal truths from physical truths exposed by the conceivability of zombies gives rise to a new issue: a failure in any reductive explanation of phenomenal truths in terms of physical truths. The fact that we can conceive of the instantiation of the physical property without the phenomenal one shows an *explanatory gap* (Levine 1983). In order to derive the explanatory gap, the only thing we need to assume is a link between conceivability and the reductive explanation we are looking for; the kind of link that makes it transparent why some high-level truth obtains whenever some low-level truths obtain; i.e. an a priori entailment of higher-level truths from lower-level truths.

Whereas 'water is H_2O' is explanatory it seems that 'being in pain is having such-and-such neural activity' leaves something unexplained. There is a certain entity that is odorless, colorless and fills rivers and lakes, we call it water. Our knowledge of physics and chemistry explains how H_2O causes the manifested properties we associate with water: H_2O is what plays the causal role of water –nothing is left unexplained once we have understood how this causal role is fulfilled by water. In the case of pain one could maintain that part of the concept of pain is also captured by a causal role. When one is in pain, one tends to have the belief that something is wrong with the body, the desire to be out of that state and to say 'ouch'. Pain alerts us of damage and helps us to avoid certain situations related to damage, etc. If I burn my finger, such-and-such neural network is activated by the heat, which will trigger certain beliefs, desires, and behavioral responses characteristic of being in pain. So, such-and-such neural network could be a good candidate for being the realizer of this causal role. However, there seems to be more to phenomenally conscious states than satisfying this causal role. There is something it is like to be in a phenomenally conscious mental state, there is something it is like to be in pain: it *feels* a certain way. The interesting question for phenomenal consciousness is left unexplained: why does having such-and-such neural activity *feel* at all, and in such a particularly characteristic way?

The knowledge argument exploits the lack of a priori entailment between phenomenal truths and physical truths to show a problem in the materialist's explanation. Some philosophers have argued that the right conclusion to be derived from this failure in the reductive explanation of phenomenal truths from physical truths is that there is an ontological gap between phenomenal properties and physical properties.

In subsection 2.2.1, I will make some considerations about the knowledge argument. In the first place, I will discuss whether reductive materialism is in a worse position than reductive dualism. Surely, if we look for a satisfactory theory of phenomenal consciousness this reply is not enough. We still need to explain, in a way that is compatible with materialism, the reasons of the failure of a reductive explanation of phenomenal properties. In second place, I will consider some reasons given by Brown for doubting about the conclusion of the knowledge argument. I will maintain that Brown's argument is interesting but insufficient for rejecting that there is a problem for materialist theories behind the argument.

Advocates of the explanatory gap hold the following thesis:

> (A priori entailment thesis)
>
> If L reductively explains H, then it is a priori that $L \rightarrow H$.

In subsection 2.2.2 I will discuss the a priori entailment thesis in some detail and three different views on the relation between a priori entailment and reductive explanation. I will conclude that the lack of a priori entailment shows that there is a failure in the explanation exclusive of phenomenal consciousness (or at least not ubiquitous in scientific explanation) and deny that the right conclusion to be derived from this gap is the truth of dualism. To block this conclusion an explanation of the failure in the a priori entailment has to be provided: this is the task of the phenomenal concept strategy in the last section.

2.2.1 *Some Considerations about the Knowledge Argument*

In this section I will be discussing some of the reasons that have been given for doubting that the Knowledge Argument against materialism is convincing.

The Knowledge Argument Against Dualism

Paul Churchland (1989) has argued that the Knowledge Argument goes too far. If it is sound, it shows not only that materialism is false, but also that dualism (or at least substance dualism) is false. According to substance dualism, there is another substance different from the physical in virtue of which phenomenally conscious experiences have phenomenal character. It seems equally plausible that knowledge about such non-physical substance doesn't give us knowledge about the phenomenal character of conscious experiences. Churchland claims that:

> Given Jackson's antiphysicalist intentions, it is at least an irony that the same form of argument should incidentally serve to blow substance dualism out of the water (Churchland, 1989, p. 574).

Following this idea, Nagasawa (2002) presents an argument in which Mark, a future scientist, comes to know all physical and non-physical facts about the world. In this future science a stuff called 'X' is discovered (or revealed) to ground all mental phenomena. Mark has all of Mary's knowledge plus all the knowledge about X. Similarly to Jackson's character, Mark obtained all of this knowledge in black and white surroundings. Nagasawa has the intuition that Mark will learn

something when he sees red for the first time and that the situation is not very different from Mary's case. If this intuition is right, Nagasawa claims, it is unfair to emphasize only the anti-materialist side of the Knowledge Argument.

One quick reply the dualist can make, if she concedes the intuition, is to acknowledge that substance dualism is as vulnerable to the Knowledge Argument as materialism, but to insist that the interesting form of dualism in play is property dualism. According to property dualism, mental states are physical states with special mental properties, properties that differ from physical properties. This quick reply seems insufficient for handling the objection. The way the argument is presented by Nagasawa is immune to this reply: it is irrelevant to the argument whether X is a property or a substance.

In order to discern the kind of dualism that would be immune to the Knowledge Argument from one that wouldn't, Nagasawa distinguishes between reductive and non-reductive dualism. The former explains mental properties in terms of some kind of lower-level properties. The latter assumes that there is no reductive explanation possible for phenomenal properties. I see two different reasons one might have for endorsing non-reductivism: The first one is epistemological. We are cognitively closed to the nature of phenomenal consciousness. However, in this case, we are left with an unappealing dualist version of McGinn's mysterianism (McGinn 1989, see 1). The second one might be that there is nothing to be reduced: phenomenal properties are some kind of fundamental primitives of the universe. This form of dualism is immune to the Churchland/Nagasawa objection, because the knowledge argument targets only reductive explanation.

As we will see in 2.2.2, some philosophers have argued that every reductive explanation requires an a priori entailment of higher-level properties from the lower-level properties. The knowledge argument seems to show that this is not possible in the case of phenomenal consciousness. If Churchland's intuition is correct, then the reductive version of dualism is in trouble. There is nothing special in the Knowledge Argument against materialism. The worry is against reductive explanations of phenomenal properties.[25]

Chalmers (1996) offers both reductive and non-reductive versions of dualism.

> There are two ways this might go. Perhaps we might take [phenomenal] experience itself as a fundamental feature of the world, alongside space-time, spin, charge and the like. That is, certain phenomenal properties will have to be taken as basic properties. Alternatively, perhaps there is some other class of novel fundamental properties from which phenomenal properties are derived. ... [T]hese cannot be physical properties, but perhaps they are nonphysical properties of a new variety, on which phenomenal properties are logically supervenient. Such properties would be related to experience in the same way that basic physical properties are related to nonbasic properties such as temperature. We could call these properties protophenomenal properties, as they are not themselves phenomenal but together they can yield the phenomenal. (Chalmers, 1996, pp. 126-127)

25 A non-reductivist materialist position like Davidson's Anomalous Monism (Davidson 1970) would be immune to the objection.

The first proposal is a form of non-reductive dualism in which phenomenal properties belong to our fundamental ontology. The latter is a form of neutral monism (Chalmers calls it 'Russellian Monism' or 'type-F materialism'), according to which phenomenal properties are reducible to proto-phenomenal properties. But one cannot rest on the Knowledge argument to argue in favor of this position, if the Churchland/Nagasawa intuition is right, for the Knowledge Argument is an argument against any kind of reduction of phenomenal properties. Whatever trick the advocate of reductive dualism appeals to in order to reply to the argument, the same trick can be reproduced by the reductive materialist.[26]

If Churchland and Nagasawa are successful in their arguments, then they show that reductive dualism is in no better position than reductive materialism with regard to the explanatory gap. However, I do not have a clear intuition about reductive dualism and I fail to see what would justify intuitions about properties we don't know. So my reply to the knowledge argument will not rest on the Churchland/Nagasawa argument. Furthermore, this argument does not, by itself, offer a satisfactory reply to the knowledge argument. To reject a connection between the explanatory gap and an ontological gap, materialists have to explain, in terms compatible with materialism, why there is a failure in reductive explanation. I will offer it in section 2.3. Let me first review further arguments for rejecting the explanatory gap.

The Irrelevance of the Knowledge Argument.

Richard Brown (2010) presented a variation of Jackson's argument to dispute the intuition supporting the Knowledge argument. In Brown's argument, the evil scientist that raised Mary as a super-scientist, raised Maria as a super-phenomenologist.

> [Maria] was raised in a special room where she was taught from a very early age to focus on her own experience. She learns to master all of the platitudes of folk psychology and so is a master of such things as that red is more like pink than it is like blue and that turquoise is more like blue than it is like red and on and on to a degree that we can only dream of. Maria is able to discriminate between shades of color that we cannot (though perhaps we could with the proper training) also she is able to describe her experience as accurately as humanly possible. She, in short, knows everything there is to know about her own experience. (ibid., p.6)

Maria doesn't know anything about science; she doesn't even know that she has a brain. One day, Mary teaches Maria all she knows. Brown's intuition is that Maria will be able to map her experiences within the physical description of the processes given by Mary.

> I have the intuition that Maria will then learn that her visual experience of red is just a brain state, just as she learns that water is H_2O. She will learn that her color experience is a physical event in her brain. Maria will learn something that she would express by saying 'oh, so that's what my color experience is!' Once she sees the identities she will be in a

26 See Nagasawa (2002) for a discussion of possible replies.

> position to deduce the qualitative facts from the physical facts a priori. (ibid., pp. 6-7)

I agree with Brown's intuition that Maria will be able to map phenomenally conscious experiences and certain brain states, and that she will learn something that she would express by saying 'oh, so that's what my color experience is!'. However, Brown derives from this example the conclusion that Maria will be in a position to deduce *a priori* the phenomenal facts from the physical facts.[27] I disagree; Maria will be able to *see* the correlation between some phenomenal properties and some physical properties. But *seeing* this correlation is not enough for being able to deduce a priori phenomenal truths from physical truths. She is just making an inference to the best explantion. Let me elaborate.

We can walk for a while with Brown and grant him that Maria can infer which phenomenal fact will obtain if certain physical facts obtain. [28] However, all that she can know is that if someone has, say, brain state B, then one has an experience as of red. We can grant that Maria will be able to infer phenomenal facts about other people from physical facts about their brains, but just in case they satisfy exactly the same description as she does. Consequently, Maria has no way to decide whether Marc, whose brain state differs from hers, say, in one neuron, instantiates the same phenomenal property as she does. Brown's story doesn't seem to support the idea that she will know which one of the properties of B are essential for having an experience as of red. So, Maria is not in a position to deduce a priori phenomenal truths from physical truths.

I am going to illustrate this epistemic problem that we already saw in Block's harder problem (2.1.2) with a new mental thought experiment. I will try to show that Maria is not in a position to deduce phenomenal facts from physical facts a priori.

In the thirtieth century, our science has improved incredibly. Among the many new devices of the day, two are relevant for the study of consciousness: the P-reader and the brain-configurator. The P-reader is a mega-computer containing all of the microphysical information about the brains of the inhabitants of the world. The brain-configurator makes it possible to configure someone's brain from a computer, namely to activate and deactivate neurons, alter their frequency of firing, etc. In microseconds the brain configurator can even operate on the subject and give her new neurons.

In the thirtieth century, scientists have learned to control biological aging and Maria has been chosen so as not to suffer senescence, but she must complete the map of all possible human experiences. With the help of the P-reader she knows all the microphysical facts about anyone's brains. She can *see*, as Brown notes, the match between her own experiences and the microphysical facts. With the help of the brain-configurator she can undergo anyone's experience, Maria can know what it is like for me to write this dissertation. With centuries of research, Maria maps every single experience into a certain brain

27 In order for Maria to acquire the kind of knowledge required she would need to have cognitive access to all of her phenomenal states. As I showed in 1.2.1 the relation between phenomenal properties and the cognitive access we have to them is controversial. I will ignore this important issue here, since nothing of what I want to argue rests on it. For the sake of the discussion I will grant that we have this kind of access.

28 Some materialist positions deny that the brain is the supervenience basis of the experience. I am using the brain for simplicity's sake. If the supervenience basis is X, then Maria can deduce the phenomenal facts from X.

configuration. Brown would claim that Maria can a priori deduce phenomenal facts from physical facts but she would not as we are about to see.

Even if this science fiction story were possible and we accepted that Maria could, in some sense, deduce phenomenal facts from microphysical facts about the brain, the *a priori* entailment thesis would remain unsatisfied. Maria cannot a priori deduce phenomenal facts from physical facts, Maria can know what any human being is feeling just by knowing facts about their brains, but Maria cannot know what someone who has a different brain from ours is feeling or whether she is feeling anything at all. Maria won't be able to decide whether a being functionally equivalent to her but with a different physical implementation, like Commander Data, has phenomenally conscious experiences or he has not. She has no way of deciding which properties are essential to consciousness. She perfectly finds neural correlates of sensations but she cannot come to know which properties of this correlate are essential to phenomenal consciousness. She has no way of deciding whether Commander Data has the properties that are essential to phenomenal consciousness or not.

The problem is serious, but I think that this is not exclusive of the materialist position. Dualism has exactly the same problem. Remember, the dualist can either be a reductivist or a non-reductivist. In the second case there is no inter-level explanation. The question remains for the reductivist: how can the reductivist dualist decide the structural level at which proto-phenomenal properties give rise to phenomenal consciousness?

In the case of water, one can claim that, ideally at least, facts about H_2O explain facts about water. One can deduce the properties of water from the properties of H_2O. Things are different in the case of phenomenal consciousness. Even if Maria's knowledge goes as far as it is possible, she cannot deduce phenomenal facts from physical facts. No matter how ideal our knowledge of the physical facts is and no matter how ideal our knowledge of the phenomenal facts is. Phenomenal concepts are isolated from other concepts. This idea is exploited by advocates of the phenomenal concept strategy to show how the explanatory gap is compatible with the truth of materialism. Before presenting such a strategy, I will discuss different reactions to the explanatory gap.

2.2.2 *Three Different Reactions to the Explanatory Gap*

A failure in the a priori entailment between physical facts and phenomenal facts is the basis of the intuition in Jackson's thought experiment and in the conceivability of zombies. We can build such an a priori deduction in the case of water but not in the case of phenomenal consciousness (Tye 1999).

Take 'F' to be a physical/functional predicate like 'substance that fills up rivers and lakes, falling from sky,...'. Then (1) is an a priori truth

(1) Water = the F (or an F)

(1) is an priori truth, according to Tye, because the reference of the term 'water' is fixed trough this description (or whatever you prefer to be F).

The complex property picked up by 'water' is associated with H_2O in the actual world. It is an empirical truth that:

(2) H_2O= the F

It is an empirical truth that:

(3) There is H_2O in place P

---

(4) There is water in place P

This deduction seems to be correct.[29] Intuitively, there is no corresponding deduction in the case of phenomenal properties. Compare the former deduction with the following, where G is also a physical/-functional predicate:

(1′) Sensation as of red = the G

(2′) Neural oscillation N in visual cortex = the G

(3′) There is Neural oscillation N in visual cortex in subject S.

(4′) There is sensation as of red in subject S.

According to Tye, the difference between the first and the second argument is that (1′) contrary to (1) is not a priori. It seems plausible that the reference of the term 'water' is fixed via the description given by F. In the case of phenomenal truths there is no such a physical/functional predicate that helps fixing the reference and therefore (1′) is not a priori. If this is true, then no amount of a priori reflection on phenomenal concepts alone will reveal a connection between physical truths and phenomenal truths, even of a contingent kind.

We can distinguish three different reactions to the explanatory gap. I will call them Dualism, Epistemologicalism and Deflactionism.

The conceivability of zombies or Mary's tale shows that there is no a prori entailment between a physical description of the world and phenomenal consciousness. Dualism and Epistemologicalism suggest a connection between conceivability and reductive explanation. They maintain that in a reductive explanation, the explanans a priori entails the explanandum. They endorse, whereas Deflationism denies, the a priori entailment thesis.

> A priori entailment thesis:
>
> If L reductively explains H, then it is a priori that $L \rightarrow H$.

Dualism (Chalmers 1996; Chalmers and Jackson 2001; Chalmers 2009) maintains that a failure in the explanation supports the conclusion that phenomenal facts do not metaphysically supervene on physical facts. For the dualist, the intuition has a metaphysical consequence: the entailment of metaphysical possibility from conceivability is the simplest and most satisfying account of the failure in the explanation. There is no a priori entailment between physical truths and phenomenal truths because phenomenal truths do not metaphysically supervene on the physical.

Epistemologicalism (Levine 1983) accepts that the intuition shows a failure in the explanation, an explanatory gap, but resists the inference from this gap to an ontological one. The intuition has epistemological consequences, but not ontological ones: the intuition is not about what is metaphysically possible.

29 As we are about to see, this is not uncontroversial.

Finally, deflationism (Block and Stalnaker 1999; Papineau 2002) denies that there is anything special about phenomenal consciousness. The failure of *a priori* entailment from truths at a lower-level to truths at a higher-level is not exclusive of consciousness. The epistemic gap between levels is sometimes closed without a priori conceptual analysis.

The positions can be summed up in the following chart:[30]

	Metaphysical Gap	Explanatory Gap
Dualism	Yes	Yes
Epistemologicalism	No	Yes
Deflationism	No	No

I will argue in favor of epistemologicalism. In this subsection I will focus on the debate between those who accept the explanatory gap, dualists and epistemologicalists, on the one hand, and deflationists on the other. I will argue that the problem cannot be completely deflated and that the epistemic gap remains. However, I will maintain that an ontological gap is not the right conclusion to be derived from the explanatory gap. In the last section of this chapter I will argue that the explanatory gap does not present a problem for materialism.

The most important point for the discussion is whether ordinary macroscopic truths about the world, such as 'there is water in this glass', are *a priori* entailed by microphysical truths, for instance truths about strings. The question is very important if one wants to hold on a thesis about reductive explanation that maintains that reductive explanation necessitates a priori entailment.

The point of discussion has to be refined. Chalmers and Jackson (2001) do not maintain that macroscopic truths are entailed merely by microphysical truths (P), but that they are entailed by microphysical truths (P) plus a 'That's All' clause (T), indexical information (I) and phenomenological truths (Q).

The 'That's All' clause is required at the end of the microphysical description to guarantee that the description is exhaustive, that there is no additional entity. The truth of 'There are no angels' cannot be entailed by P. A 'That's All' clause is additionally required to that effect.

Indexical information like 'I am here' is necessary to be able to extract context dependent information about the world.

Phenomenal truths are required for macroscopic truths that depend on them.[31] Arguably, in many cases an entity falls under the extension of a concept depending on what it looks like to us. For instance, if it were part of our concept of *water* that it is odorless and tasteless, then phenomenological information is required for having the concept *water*.

The case in favor of the explanatory gap can be summarized as follows (Diaz-Leon 2010b):

> A Priori Entailment Thesis:
>
> Reductive explanation necessitates that the explanandum is a priori entailed by the explanans.

(1) A reductive explanation of macroscopic truths in microphysical terms is possible only if macroscopic truths are a priori entailed by PQTI.

30 More precisely, the deflationist claims that the explanatory gap is ubiquitous in nature.

31 If phenomenal truths were a priori entailed by microphysical truths then this clause would be redundant.

(1′) A reductive explanation of phenomenal truths in microphysical terms is possible only if phenomenal truths are a priori entailed by PTI.

(2) All macroscopic truths are a priori entailed by PQTI.

(2′) Phenomenal truths are not a priori entailed by PTI.

---

(3) Reductive explanation of phenomenal truths in terms of physical truths is not possible.

The comparison with reductive explanation of macroscopic truths is relevant for making the case in favor of the a priori entailment thesis. Science offers us reductive explanations of macroscopic truths in microphysical terms.

Dualists and epistemologicalists hold that, given a microphysical description, an ideal competent speaker, who possess the relevant concepts, should be able, just by looking at this description and reflecting on it, to tell what ordinary facts would obtain if the world satisfied this description. This kind of link is not possible for phenomenal consciousness if we accept the ideal negative conceivability of zombies. Deflationists hold that this kind of link is not always possible for all macroscopic truths either.

Block and Stalnaker (1999) reject (1), they maintain that bridging the gap between descriptions at the level of the explanandum and descriptions at the level of the explanans cannot be done by mere conceptual analysis in the case of most a posteriori necessities. If there are *bona fide* cases of reductive explanation of macroscopic truths for which there is no a priori entailment from the explanans to the explanandum, then the a priori entailment thesis would be false. The claim that there is a failure in the explanation in the case of phenomenal truths because of the lack of a priori entailment between phenomenal truths and physical truths would be unsupported.

Suppose we want to explain the ordinary fact that water boils at 100°C. We start by conceptual analysis of the terms involved: 'water', 'boil', etc. There will be an ultimate link between the languages of the two different levels (microphysics, physics, chemistry, etc), one based just on conceptual analysis. When this link is established, one can understand that when certain facts about the *explanans*[32] obtain, certain facts about the *explanandum* will obtain. This is an a priori entailment.

Block and Stalnaker intend to show that there is no asymmetry between the explanation of phenomenal consciousness and other physical facts. They deny that lower-level truths explain higher-level truths only if one can derive, by conceptual analysis, higher-level truths from lower-level truths. Typically, there is no explicit analysis of macroscopic concepts that supports an *a priori* entailment from microphysical to macroscopic truths. For Block and Stalnaker there is not always connection between levels, this is only plausible in certain concrete cases when *explanans* and *explanandum* "involve the same 'family' of terms".

Deflationism denies that scientific explanation in general satisfies the *a priori* entailment thesis. Deflationists claim that this requirement is too high and no scientific explanation requires it. Unless the only epistemic access we have to identity depends on our semantics, it doesn't follow

32 I use facts in such a way that they include laws.

that explanation is exclusively a matter of conceptual analysis plus *a priori* entailment. The gap is filled by correlation. We look for correlation and argue from correlation to identity between levels (a posteriori) by inference to the best explanation.

Consider the following very rough explanation of why water boils when it is heated: at ordinary low temperatures, water, left open to the air, gently evaporates from its surface. The pressure of this vapor at such low temperatures is much less than the pressure of the surrounding atmosphere. The water also tries, through the formation of microscopic bubbles, to evaporate in its interior. However, this tiny bubbles of water vapor are immediately suppressed, because the pressure of the atmosphere pressing down on the liquid's surface is much higher than the pressure of the vapor. *The kinetic energy of the* H_2O *molecules increases, when water is heated*. This causes more and more molecules to escape the liquid increasing the pressure. When the vapor pressure reaches the pressure of the surrounding air, the bubbles that form by evaporation in the interior of the liquid are no longer suppressed and water starts to boil. If heat is identical to molecular kinetic energy, then we have an explanation of why water boils when it is heated.

> Identities allow a transfer of explanatory and causal force not allowed by mere correlations [...] Thus, we are justified by the principle of inference to the best explanation in inferring that these identities are true. Block and Stalnaker (1999)

The main problem for the advocate of the explanatory gap is that it is not clear that conceptual analysis is always possible in the case of macroscopical truths as required by the a priori entailment thesis. Block and Stalnaker consider the case of life, where no a priori conceptual analysis is required for reduction, for it seems "doubtful that fulfilling any set of functions is conceptually sufficient for life" (Block and Stalnaker, 1999, p. 377).[33]

This problem is addressed, correctly I think, by Chalmers and Jackson who deny that this kind of explicit analysis is required by the a priori entailment thesis:

> Once an essential role for explicit definitions is eschewed, the model of conceptual analysis that emerges is something like the following. When given sufficient information about a hypothetical scenario, subjects are frequently in a position to identify the extension of a given concept, on reflection, under the hypothesis that the scenario in question obtains ... What emerges as a result of this process may or may not be an explicit definition, but it will at least give useful information about the features in virtue of which a concept applies to the world ... The possibility of this sort of analysis is grounded in the following general feature of our concepts.

33 Block and Stalnaker offer further arguments. They claim that functional analysis requires a *uniqueness* thesis (p.379): water should be analyzed as something like 'the unique waterish stuff in our environment'. They argue that it could be the case that there is more than one kind of stuff in the environment that satisfies the relevant descriptions. But Chalmers and Jackson deny both, that *a priori* entailment requires explicit analysis and that it is required to analyze water as something like 'the unique waterish stuff in our environment'. This and other reasons provided by Block and Stalnaker are convincingly refuted by Chalmers and Jackson (2001, section 5).

> If a subject possesses a concept and has unimpaired rational processes, then sufficient empirical information about the actual world puts a subject in a position to identify the concept's extension. (Chalmers and Jackson, 2001, pp. 322-323)

Another objection against Chalmers and Jackson's position has been presented by Diaz-Leon. She argues that full grasp of the meaning of the terms involved in the conditional required for the explanation of macroscopical facts in microphysical terms requires empirical knowledge and therefore the conditional is not a priori.

Diaz-Leon (2010b) recalls the distinction between full grasp and deferential grasp of the meaning of a term. For most of our concepts, we are deferential users of them; we do not need to know exactly under which conditions an entity falls under the extension of a concept. Rather, we rely on experts. Fortunately, full knowledge of the conditions under which an entity falls under the extension of a concept is not required for concept possession, but it is indeed required by the a priori entailment thesis.

Arguably, a non-deferential user of a concept has a full grasp of the meaning of such a concept; given a sufficiently rich description of an scenario, she knows what determines whether an entity falls under the extension of a concept. Full grasp requires a conditional ability to apply the concept to the entity the concept refers to when it is described at a lower level. But the mechanisms in virtue of which an expert acquires full grasp of a concept are paradigmatically empirical. Diaz-Leon argues that if knowledge of the *explanans* is not *a priori*, then the knowledge of the conditional involved in the explanation can only be said to be *a priori* in a technical sense.

Chalmers (2010, p. 220; fn. 16) replies that even if Diaz-Leon is right and empirical knowledge is required for full grasp of a concept, such an empirical knowledge has only an enabling role, rather than an epistemic one justifying the hypothesis.

I agree with Chalmers. The idea that supports the a priori entailment thesis is the following: an ideal subject, possessing the relevant concepts, sat down in a sofa, would be able to deduce, given a description of a possible world using lower-level truths, which higher-level truths obtained in such a world. For instance, from a description, AD, at the level of atomic reactions, an ideal subject can come to know that if the world satisfies this description then there is water boiling in the pot.

Our ideal subject has to have full grasp of the meaning of 'water' or 'boiling'. Diaz-Leon is complaining that full possession of concepts requires empirical knowledge. However, as Chalmers rejoins, this empirical knowledge is required exclusively to be able to entertain the reductive conditional; i.e. if AD obtains then there is water boiling in the pot. The conditionals are justified independently of the experience. The justification of the conditional is based just on a priori reasoning. Once the ideal subject has the concepts required for entertaining the conditional, the empirical knowledge plays no role in the justification. She can justify the conditional sat down in her sofa. On the other hand, in the case of phenomenal consciousness such an ideal subject would not be able justify a priori the conditional from physical truths to phenomenal truths, even if she had full grasp of both physical and phenomenal concepts, as the conceivability of zombies or Mary's story intends to show.

For that reason, I will assume that there is an explanatory gap; contrary to other inter level conditionals there is no a priori entailment between physical truths and phenomenal truths. Materialism has to provide either a way to close the gap or the second best thing: a materialistic explanation of why there is or seems to be an explanatory gap. The phenomenal concept strategy offers such an explanation.

2.3 PHENOMENAL CONCEPT STRATEGY

I have finished sections 2.2 and 2.3 with a half reply to the anti-materialist arguments. In order to block the modal argument, we have seen that we only need to show that anti-zombies are negatively conceivable; i.e. that there is no logical contradiction in the idea of a minimal physical duplicate of me that undergoes phenomenally conscious experiences. Furthermore, we need to explain, in terms that are compatible with materialism, the conceivability of zombies.

In addition, I have conceded that there is an explanatory gap between physical truths and phenomenal truths. A full-blown reply to the anti-materialist argument would deny the entailment from the explanatory gap to an ontological gap by explaining in terms compatible with materialism how is it possible that there is no a priori entailment from physical truths to phenomenal truths despite the fact that phenomenal properties metaphysically supervene on physical properties. The phenomenal concept strategy attempts to provide an answer to these issues.

The so-called phenomenal concept strategy claims that phenomenal concepts are special and that some anti-materialist arguments take their force from a misunderstanding of their special nature. The idea supporting the phenomenal concept strategy is that phenomenal concepts differ importantly from physical concepts. According to the advocate of the phenomenal concept strategy, the perplexity put forward by the explanatory gap and other anti-materialist arguments rests on an ignorance of this difference.

Phenomenal concepts are the concepts we deploy to refer to experiences with certain phenomenal character.

Daniel Stoljar (2005), who introduced the name of 'phenomenal concept strategy', presents phenomenal concepts as follows:

> A phenomenal concept is the concept of a specific type of perceptual or sensory experience where the notion of experience is understood phenomenologically. So, for example, the phenomenal concept RED SENSATION is the concept of the specific type of sensation one gets from looking at red things such as British pillar-boxes or the Chinese flag. The concept RED SENSATION is not then the concept RED, for that concept typically qualifies objects not sensations. Nor is it the concept SENSATION THAT REPRESENTS THINGS AS RED, for there is no contradiction in the idea of a red sensation that did not represent things that way. Nor even is it the concept THE SENSATION ONE GETS FROM LOOKING AT RED THINGS, for that sensation might not have been a red sensation.

I will use '#PC_X' to refer to the phenomenal concept of experiences with phenomenal character PC_X. The phenomenal concept #PC_{RED}

refers to an experience as of red, experiences in which there is a *redness way it is like for someone* to undergo them.

I do not think that there is such a thing as an experience as of red in general, but maximally concrete *redness ways it is like for me* to have an experience. There is no such thing as an experience of seeing something red in general; instead we have different *redness* ways: the way it is like for me to look at the red apple under certain concrete lighting conditions, to look at the sunset on a autumn evening, or to my wound bleeding in my room. $\#PC_{RED_1}$, $\#PC_{RED_2}$,.... refer to these different ways that we identify as experiences as of red. The phenomenal concept $\#PC_{RED}$ refers to experiences with one of these phenomenal properties.

We can therefore distinguish between specific and general phenomenal concepts. I will call 'Specific Phenomenal Concepts' (SPC) the concepts that refer to experiences like the one I am having right now while looking at my red apple (I am not at all maintaining that I cannot undergo another experience with the same phenomenal character, nor that another subject cannot). I will call 'General Phenomenal Concepts' (GPC) the concepts that refer to experiences with phenomenal characters similar in some sense, as shown in the introduction. Examples of GPC are $\#PC_{VISUAL}$, which refers to visual experiences in general, $\#PC_{RED}$, which refers to the experiences I typically have when I look at a red object in general or $\#PC_{FEEL}$, which refers to any phenomenally conscious experience.

In 'Phenomenal States' Brian Loar (1990) suggests the idea that phenomenal concepts are different from other physical concepts, that they do not work in the same way. According to Loar, phenomenal concepts are direct recognitional concepts. When we are having an experience we can deploy a concept that refers directly to the phenomenal character of this experience. Loar's ideas also suggest that phenomenal concepts involve the experience itself.

Current theories about phenomenal concepts can be grouped into two depending on which of these ideas is developed in the theory (Balog 2009). On the one hand, direct reference accounts (Aydede and Guzeldere 2005; Perry 2001; Tye 2003b) focus on the fact that phenomenal concepts pick out their reference directly, there is a direct relation between phenomenal concepts and phenomenally conscious experiences. On the other hand, special modes of presentation (Balog mingc; Carruthers 2003; Block 2006; Hill and Mclaughlin 1999; Papineau 2002; and Chalmers 2003b on the dualist side) accounts intend to capture the special intimacy between phenomenal concepts and phenomenal characters by suggesting that the mode of presentation of a phenomenal concept involves the phenomenally conscious experience itself.[34]

My aim in this chapter is not to discuss the differences among theories of phenomenal concepts, but instead defend the materialist strategy based on the special nature of these concepts. For my purposes it is reasonable to see advocates of the phenomenal concept strategy as defending a view concerning phenomenal concept possession that Stoljar calls the experience thesis:

EXPERIENCE-THESIS: S possesses the (phenomenal) concept #PC of experience E only if S has actually had experience E.

The phenomenal concept strategy suggests that phenomenal concepts are very different from other concepts. It maintains that, due to certain

34 For a detailed taxonomy see (Balog, 2009).

features of phenomenal concepts, say the experience-thesis, physical concepts and phenomenal concepts are not a priori linked. Thoughts that connect phenomenal and physical concepts are always a posteriori. The problem with psyco-physical/functional identities is therefore a confusion between the sense and the referent of the concept. We use two different concepts to refer to a unique referent. We think of a single state under two different concepts, being the phenomenal one conceptually irreducible to the physical one. There is a difference in the role that the concepts play but not in their referents.

We use phenomenal concepts for discriminating phenomenal qualities and states directly on the basis of introspection (assuming normal conditions). They help us via a reliable process to know that we are in a certain state, for instance, phenomenally seeing red. On the other hand, when we think of the very same referent under the theoretical physical concept, we think of it in a different way, for instance feedback loops involving among others neural networks in areas V1 and V4 of the visual cortex and oscillations in the gamma range (30-70 Hz). Crick and Koch (1990).

The special nature of phenomenal concepts explains why no a priori connection can be found between phenomenal facts and physical facts. Even if the referent of $\#PC_{RED}$ is a physical property, for example the very same physical property as the one referred to by the expression 'having neural oscillation N in the visual cortex', Mary cannot come to know that having neural oscillation N in the visual cortex is having an experience with phenomenal character PC_{RED} in an intensional sense (the sense in which knowing that Hesperus is a planet does not entail that one knows that Phosphorus is a planet). She cannot come to know what it is like to see red, for she lacks the phenomenal concept $\#PC_{RED}$ required for it. She has never undergone such an experience. Lacking the phenomenal concept she has no idea of how it *feels* to see a red rose. Mary knows all the physical and functional facts about color vision. She knows that when someone sees a red apple, he is in a certain brain state; she cannot, however, deduce from that how being in this state feels. The phenomenal is not deducible from the physical. Chalmers and Jackson's demand cannot be satisfied and the phenomenal concept strategy presents an explanation compatible with the truth of materialism of why this is so.

Given the nature of phenomenal concepts, there is no connection between the phenomenal and physical concepts. This explains why, even if I had the corresponding phenomenal concept $\#PC_{RED}$, I knew all the physical facts about color vision and I knew that you are in the corresponding physical/functional state that $\#PC_{RED}$ refers to, I could not build an a priori connection between my concept and your state. I cannot a priori know what you *feel* and I could conceive of you being in that state and having a yellow sensation.

As we saw in the previous section, we can build such an a priori deduction for physical terms, like water, but it fails in the case of phenomenal consciousness. Recall the two deductions presented on page 59:

(1) Water = the F (or an F)

(2) H_2O= the F

(3) There is H_2O in place P

∴ There is water in place P

And in the case of phenomenal consciousness

(1′) Having a PC_{RED} experience = the G

(2′) Neural oscillation N in visual cortex = the G

(3′) There is Neural oscillation N in visual cortex in subject S.

∴ S has an experience with phenomenal character PC_{RED}.

The second argument is not a priori sound. The reason for that is, following Tye (1999), that (1′) is not a priori true. The phenomenal concept strategy maintains that due to certain features of phenomenal concepts, physical concepts and phenomenal concepts are not a priori linked. Thoughts that connect phenomenal and physical concepts are always a posteriori.

For instance, Hill and Mclaughlin (1999) suggest that the concepts have different 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 a 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, p. 453)

It is part of the functional role of phenomenal concepts that they enable us to discriminate between phenomenal qualities and states directly on the basis of introspection, without descriptive reference fixing intermediaries. The reference of phenomenal concepts is not fixed via any description. We cannot see a priori that physical and phenomenal concepts are co-referential. No amount of a priori reflection on phenomenal concepts alone will reveal phenomenal-physical or phenomenal-functional connections, even of a contingent type.

This is the general structure of the phenomenal concept strategy. In the next section I will discuss two objections to the phenomenal concept strategy. In the first place, Tye (2009) claims that the concepts that refer to phenomenal characters are not relevantly different from other concepts. If we understand phenomenal concepts as concepts with a special nature, Tye denies that there are such phenomenal concepts. In the second place, Chalmers (2007) maintains that either the nature of phenomenal concepts cannot be explained in a way compatible with materialism or, if it can, then what cannot be explained is our

epistemic situation[35] with regard to the explanatory gap. In either case materialism is jeopardized.

2.3.1 *Objections to the Phenomenal Concept Strategy*

There are no Phenomenal Concepts

In his last book 'Consciousness Revisited: materialism without phenomenal concepts' (Tye, 2009), Michael Tye rejects his former views on phenomenal concepts (Tye, 1999, 2003b), and maintains that phenomenal concepts, the concepts we deploy when we entertain any thought about the phenomenal character of our experience, are not different from other concepts in any relevant sense that helps saving materialism.

As we have seen, it is widely suggested among philosophers that argue in favor of the phenomenal concept strategy, that the mechanisms needed to possess a phenomenal concept include in some respect or other phenomenally conscious experiences. This is what Stoljar (2005) calls the experience thesis:

EXPERIENCE-THESIS: S possesses the (phenomenal) concept #C of experience E only if S has actually had experience E.

Tye argues that one can have the kind of understanding required for possessing the relevant phenomenal concepts without undergoing or having undergone the corresponding experience. A partial understanding is sufficient for possessing the phenomenal concept and partial understanding does not require the experience.

> Maybe *fully* understanding a general phenomenal concept requires having the relevant experience; but if such concepts are like most other concepts, possessing them does not require *full* understanding. They can be possessed even if only partially understood. (Tye, 2009, p. 63; emphasis in the original)

Tye seems to be considering the idea, emphasized by Putnam (1975) and Burge (1979), that one can possess, say, the concept ELM without knowing much about elms. Those who possess this concept are typically willing to accept correction from others about its extension. As Tye puts it, the concept is deferentially used. Tye holds that the same is true about phenomenal concepts.

In Burge's example, a patient believes he has developed arthritis in his thigh. When his doctor explains him that arthritis cannot occur in the thigh, because arthritis is a disease of the joints, the patient accepts that his earlier belief was false. As Tye emphasizes, the possibility of such agreement seems to require that they share a single concept.

Tye argues that similar reasoning applies to phenomenal concepts. According to him, phenomenal concepts can also be deferentially used. For example, someone undergoing dental work might, at first, classify her experience as pain but later accept corrections from an expert who maintains that the experience was actually a case of pressure. Furthermore, Mary before being released might share various beliefs about the phenomenal character of color experiences with colorsighted people outside the room. For example, she might agree that seeing red is phenomenally more similar to seeing orange than to seeing

35 The notion of epistemic situation is technical; I will present it properly in the discussion.

green. Tye concludes that our phenomenal concepts are not *experientially perspectival*: possessing them does not require undergoing relevant experiences. This result would undermine the phenomenal concept strategy. Tye's argument against the phenomenal concept strategy can be summarized as follows:

(1) If a concept C can be deferentially used, then partial understanding of a concept C is sufficient for possessing C.

(2) Every phenomenal concept can be deferentially used.

(3) It is not necessary to undergo the relevant experience to have partial understanding of a phenomenal concept.

∴ It is not necessary to undergo the relevant experience for possessing any phenomenal concept.

Tye maintains that phenomenal concepts can be partially understood without having the relevant experience. For instance, someone who has never experienced red can know that the phenomenal character of an experience as of red is more similar to the phenomenal character of an experience as of orange than to an experience as of green or that fire engines typically cause experiences as of red. Tye denies that phenomenal concepts differ in any interesting sense from other ordinary concepts.

He further suggests that phenomenal concepts can be deferentially used because judgments about the phenomenal character of experience can also be corrected: we are willing to accept corrections about how to apply them in some cases. One can accept correction of her thought that the color of the walls is clearly red if all her friends agree that it is a borderline case between red and orange. If one is willing to accept correction with regard to colors, one should be willing to accept correction as to whether her experience should properly be counted as having one phenomenal character or other.

I am going to grant to Tye that the concepts we deploy for referring to some color experiences can be deferentially used, despite the fact that I do not find his support of this claim appealing at all. I think that even if one concedes that the experiential thesis is not strictly speaking true, one can show that Tye's argument does not undermine the phenomenal concept strategy. Let me elaborate.

We have seen in the discussion between Diaz-Leon and Chalmers on page 64 that the kind of a priori entailment we are interested in requires full understanding of the concepts involved. An ideal subject sat down in her sofa has to have full understanding of the concepts involved in order to be able to know what higher-level truths would obtain if lower-level truths obtained. To be able to come to know that if the world were to satisfy a description AD at the level of atomic reactions, then there would be water boiling in the pot, one requires full grasp of the concepts involved, partial understanding does not suffice (I wouldn't be able to do it!).

The advocate of the explanatory gap maintains that even if an ideal individual had full grasp of the concepts involved, she could not derive phenomenal truths from physical truths. Tye seems to be happy to concede that full understanding of a phenomenal concept requires the experience: "Maybe *fully* understanding a general phenomenal concept

requires having the relevant experience" (Tye, 2009, p. 63). In this respect phenomenal concepts are special in nature, fully understanding the concept requires having undergone the experience.

Furthermore, even if partial understanding suffices for possessing some phenomenal concepts, it is completely implausible that all phenomenal concepts can be so possessed by an individual: not all phenomenal concepts we possess can be acquired without the relevant experience. A zombie cannot even deferentially possess any phenomenal concept.

Based on this idea, we can develop a second way of blocking Tye's argument. This argument has been proposed by Lynn:

> To acquire the semantic competence of deferential phenomenal concepts, one must also possess non deferential categorical concepts that are also phenomenal. Thus not all phenomenal concepts can be possessed without the relevant experiences. A philosophical zombie can never acquire any deferential phenomenal concepts.

One can possess a concept either with full understanding or with partial understanding. In order to partially understand a concept, it is required that one is competent in deferentially using the concept. This requires certain knowledge about the kind of experts one can rely on. In the Burge's famous example I might not fully understand the concept ARTHRITIS but we would not consider that I am competent in deferentially using this concept if I think that the expert I have to rely on is my tailor and not my doctor. We would not say that I am competent in deferentially using the concept ARTHRITIS unless I know that it is a disease. In order to have partial understanding of the concept ARTHRITIS one must possess the concept DISEASE. The concept DISEASE is a categorical concept for the concept ARTHRITIS.

Let's consider another example. Even if we can accept that someone who believes that a fortnight is ten days has a partial understanding of the concept FORTNIGHT, we would deny that someone who believes that a fortnight is a car model or that a fortnight is fourteen ducks would count as having a partial understanding of the concept. In order to partially understand the concept FORTNIGHT one needs to understand that a fortnight is a period of time. The concept of PERIOD OF TIME is a categorical concept for the concept FORTNIGHT.

We can now attempt to provide a definition of the notion of categorical concepts:

> A concept C_{CAT} is a categorical concept for another concept C if, and only if, if a subject S deferentially possesess the concept C then S possesess C_{CAT}.[36]

In Burge's example the concept DISEASE is a categorical concept for the concept ARTHRITIS, because anyone who deferentially possesses the concept ARTHRITIS must possess the concept DISEASE.

We can now present Lynn's argument against Tye as follows:

(1) A subject S can possess a concept C with partial understanding only if S also possesses at least one appropriate categorical concept such that the referent of the former also falls under it.

36 C_{CAT} doesn't have to be fully possessed. In this case, however, the partial understanding of C_{CAT} requires a further categorical concept for C_{CAT}.

(2) S possesses a phenomenal concept deferentially only if S also possesses at least one appropriate phenomenal categorical concept. One phenomenal concept under which the relevant experience also falls under.

This premise is just the extension of (1) to phenomenal concepts. Here the distinction I made between general phenomenal concepts (GPC) and specific phenomenal concepts (SPC) will help us to understand the premise. The categorical concept of a SPC is a GPC. One cannot deferentially possess the phenomenal concept $\#PC_{RED_{34}}$ unless one possesses a general phenomenal concept like, for instance, $\#PC_{VISUAL}$; i.e. unless one knows that it refers to a phenomenally conscious visual experience.

According to Tye, one can have a partial understanding of a phenomenal concept such as $\#PC_{RED}$ without having undergone any experience as of red. One can believe that fire engines typically cause experiences as of red without having undergone any experience as of red. However, we would not accept that one can partially understand this concept unless one posseses a GPC like, for example, $\#PC_{VISUAL}$. Otherwise, one might correctly have the belief that fire engines cause experiences as of red but also believe that the referent of $\#PC_{RED}$ "falls also under DOG-PEE, a non-phenomenal concept that refers to the property of often being urinated on by dogs."(Lynn) We would not accept that this subject possesses the phenomenal concept $\#PC_{RED}$.

(3) Phenomenal categorical concepts can be possessed either deferentially or non-deferentially.

(4) $\#PC_{FEEL}$ cannot be possessed deferentially.[37]

It does not seem plausible to maintain that a zombie can have partial understanding of the concept $\#PC_{FEEL}$, the phenomenal concept under which all phenomenally conscious experiences fall under. The reason is that $\#PC_{FEEL}$ has no categorical concept one can appeal to in order to warrant a partial understanding. The only way one can understand what is a phenomenally conscious experience is by having undergone phenomenally conscious experiences.

$\therefore$ S possesses a phenomenal concept deferentially only if S possesses a phenomenal concept non-deferentially; i.e. if S has full understanding of some phenomenal concepts.

The conclusion of the argument is that not all phenomenal concepts can be possessed without the relevant experiences. Tye's argument requires that ALL phenomenal concepts can be deferentially possessed. This is not possible. Zombies cannot possess the phenomenal concept $\#PC_{FEEL}$, they cannot even partially understand what is a phenomenally conscious experience; they cannot deferentially possess any phenomenal concept. One cannot possess any phenomenal concept unless one can undergo phenomenally conscious experiences.

The possession of phenomenal concepts with partial understanding already presupposes fully understanding other phenomenal concepts

37 I believe that GPCs do not admit deferential uses in the sense required by Tye. One cannot possess the concept #PCVISUAL unless one has undergone phenomenally conscious visual experiences. Be that as it may, the argument only requires that #PCFEEL does not admit non-deferential uses.

and full understanding of a phenomenal concept requires undergoing the relevant experience. If this is right, the phenomenal concept strategy has nothing to fear from Tye's argument.

Phenomenal Concepts are not compatible with Materialism

Chalmers (2007) presents a master argument against the phenomenal concept strategy. The argument has the form of a dilemma. If C is the key feature of phenomenal concepts responsible for our epistemic relation to phenomenal consciousness, then either C cannot be explained by the materialist or C cannot explain our epistemic situation. In either case materialism is again jeopardized.

Chalmers claims that the advocate of the phenomenal concept strategy attributes to human beings certain psychological features, C, such that:

> (i) C is true: humans actually have the key features; (ii) that C explains our epistemic situation with regard to consciousness: C explains why we are confronted with the relevant distinctive epistemic gaps; and (iii) that C itself can be explained in physical terms: one can (at least in principle) give a materialistically acceptable explanation of how it is that humans have the key features. (Chalmers, 2007, p. 172)

The phenomenal concept strategy's aim is precisely to account for our epistemic situation in such a way that is compatible with materialism, so the materialist has to satisfy (ii).

The phenomenal concept strategy has to show how materialism is compatible with the explanatory gap, so the proponent has to show how physical facts give rise to the key feature C of phenomenal concepts; they have to satisfy (iii).

Chalmers argues that no account of phenomenal concepts can simultaneously satisfy (ii) and (iii). Being P a microphysical description of the world, the argument is the following (2007, p.174).

(Anti PCS)

(1) Humans have C.

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

...

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

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

Chalmers needs some assumptions for his argument to be sound.

The notion of explanation involved in (Anti PCS) is reductive explanation. Chalmers' master argument assumes a connection between conceivability and reductive explanation. This kind of explanation requires that lower-level truths (L) a priori entail higher-level truths (H), something that we have conceded to Chalmers. The a priori entailment thesis is enough to guarantee (2) as a horn of the argument. If P&¬C

is conceivable then P → C is not a priori and therefore C cannot be reductively explained in physical terms.

The question as to whether P&¬C is conceivable is the question as to whether there is an imaginable world, which is a microphysical duplicate of the actual world but which does not satisfy the key feature of phenomenal concepts, C. If we can imagine such a world,[38] then P&¬C is conceivable. For example, the proponent of the phenomenal concept strategy maintains that Chalmers has C. If we can imagine that Chalmy does not satisfy C, then P and not C is conceivable.

The conceivability of P&¬C will depend on the details of the phenomenal concept account, but we can abstract from them in Chalmers' master argument (Anti PCS). If P&¬C is not conceivable then we can offer an explanation of the key feature of phenomenal concepts that is compatible with materialism. Unfortunately, in this case the materialist faces the other horn of the argument.

Chalmers argues that the second horn is not a more comfortable place to stay for the materialist. If our account of phenomenal concepts can explain C, then it cannot explain the epistemic gap. The reason is that if a phenomenal concept account can reductively explain in physical terms C, then C is a priori entailed by P, the microphysical description of the world. In this case even zombies satisfy C. They are microphysically identical to us. Chalmers argues that zombies do not share our epistemic situation with regard to consciousness. If he is right, then C cannot explain our epistemic situation, for C does not entail a priori (a requirement for explanation) that someone is in our epistemic situation with regard to consciousness. If we can conceive of a zombie satisfying C but not sharing our epistemic situation:[39]

(3) If P&¬C is not conceivable, then zombies satisfy C.

(4) Zombies do not share our epistemic situation.

(5) If zombies satisfy C but do not share our epistemic situation, then C cannot explain our epistemic situation.

---

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

For zombies to share our 'epistemic situation', they should be able to have beliefs. In order to have the ability to form beliefs zombies need concepts, and for that purpose, zombies should be able to have intentional states, mental states that are directed or about entities and properties in the world. The Zombies-world is microphysically identical to the actual world. Zombies have brains identical to ours, they bear the same kind of causal relation to the world and they have the same

38 As we have seen in 2.1.3, what is relevant is that an ideal conceiver can do it, for the argument requires ideal positive conceivability.

39 It has to be taken into account that these premises are bounded by a conceivability operator. The formalized version of the argument offered by Chalmers is the following:

> In a formalized version of the argument above, where E represents our epistemic situation, [4] might say that P&~E is conceivable, [3] might say that if P&~C is not conceivable, then if P&~E is conceivable, P&C&~E is conceivable, and [5] might say that if P&C&~E is conceivable, then C cannot explain E. Premise [3] is slightly more complicated in this version, in order to capture the crucial claim that the specific zombie relevant to [4] satisfies C.

history as we have. Zombies would count as having intentionality in most of the current theories of meaning (Balog (1999)).

On a Davidsonian account (Davidson 1984), zombies will have intentionality because they are just as interpretable as conscious beings. The same goes for such theories as the informational account (Dretske 1988), the causal-historical account (Kripke 1980), the counterfactual account (Fodor 1990), the teleosemantic account (Millikan 1989; Papineau 1993), the etiological account (Martinez 2010), etc. The only account in which zombies do not count as having intentionality is the account in which consciousness is required (Searle 1992), but it seems to me that no type-B materialist would endorse a theory of meaning that requires consciousness for concept possession.

In order to evaluate whether or not zombies share our epistemic situation we need to get clear about what an epistemic situation is. Let's pay attention to Chalmers' introduction of the term:

> I will take it that the epistemic situation of an individual includes the truth-values of their beliefs and the epistemic status of their beliefs (as justified or unjustified, and as substantive or insubstantive). We can say that a zombie shares this epistemic situation when the zombie has corresponding beliefs all of which have corresponding truth-value and epistemic status.
>
> Here, we assume an intuitive notion of correspondence between the beliefs of a conscious being and the beliefs (if any) of its zombie twin. For example, corresponding utterances by a conscious being and its zombie twin will express corresponding beliefs. Importantly, this notion of correspondence does not require that corresponding beliefs have the same content. It is plausible that since a zombie is not conscious, it cannot have beliefs with exactly the same content as our beliefs about consciousness. But we can nevertheless talk of the zombie's corresponding beliefs. The claim that a zombie shares a conscious being's epistemic situation requires only that it has corresponding beliefs with the same truth-value and epistemic status. The claim does not require that the zombie have beliefs with the same content. (Chalmers, 2007)

Chalmers sets two conditions for the sameness of epistemic situation:

TV	The truth values of the corresponding beliefs are the same.
ES	There is a match in the epistemic status of the corresponding beliefs.

According to Chalmers, humans and zombies do not satisfy these conditions. So, zombies do not share our epistemic situation. The conclusion of Chalmers' master argument follows: an account of phenomenal concepts either cannot be physically explainable or it cannot explain our epistemic situation with regard to consciousness.

In what follows, I am going to offer a reply to Chalmers' argument. For that purpose it is important to note that the key feature C can be conceptualized, according to the advocate of the phenomenal concept strategy, using phenomenal language (C_{phen}) and using physical language (C_{phys}). Exactly as in the case of 'having neural activity N'

and 'having a sensation as of red', there is just one property involved, conceptualized in two different and independent ways.[40] We can split into two the two premises of Chalmers' argument:

(2phys) If $P\&\neg C_{phys}$ is conceivable, then C is not physically explicable

(2phen) If $P\&\neg C_{phen}$ is conceivable, then C is not physically explicable

(6phys) If $P\&\neg C_{phys}$ is not conceivable, then C cannot explain our epistemic situation

(6phen) If $P\&\neg C_{phen}$ is not conceivable, then C cannot explain our epistemic situation

From this four premises, two of them, (2phys) and (6phen), have a false antecedent and are therefore vacuously true.

Accepting Chalmers and Jackson's thesis, any true physical description is entailed by the microphysical description of the world, P. Consequently: $P\&\neg C_{phys}$ is not conceivable and (2phys) is trivially true.

In the case of (6phen) anyone that accepts the conceivability of zombies (that $P\&\neg Q$ is conceivable) will accept the conceivability of phenomenal concept zombies, creatures that are microphysically identical to us but lacking phenomenal concepts: $P\&\neg C_{phen}$ is conceivable. The antecedent of (6phen) is false and therefore (6phen) is trivially true.

We have to focus on (2phen) and (6phys) to reply to the argument. I will argue that neither (2phen) nor (6 phys) are problematic for materialism.

On the one hand, if C has to be cast in phenomenal terms in order to explain our epistemic situation, then the phenomenal concept strategy naturally explains why C cannot be explained in physical terms (first horn). $P\&\neg C_{phen}$ is conceivable, but there is nothing new in this gap, this is precisely the very same gap that the phenomenal concept strategy was developed to account for. (2phen) is true but this is not a problem for materialism.

On the other hand, if our epistemic situation and the epistemic gap has to be cast in terms that are neutral with regard to phenomenal consciousness, then either zombies share our epistemic situation or the explanatory gap can be explained despite $P\&\neg C$ not being conceivable. That is to say, either (4) is false and hence the subargument that has (6) as conclusion is not valid or, if it is true, then it doesn't pose a problem for materialism. Let me start with this second horn.

The second horn

In this first part of the reply I will try to deny that zombies do not share our epistemic situation ((4) in the argument).

Chalmers and Chalmy share their epistemic situation if the corresponding beliefs satisfy TV and ES. According to Carruthers and Veillet (2007) they do.

In order to motivate their claim, they appeal to Chalmers' view on Putnam's twins.[41] Chalmers maintains that Oscar and Twin Oscar share their epistemic situation. When Oscar believes that water (H_2O)

40 I am following Balog (mingc) in this distinction.

41 For details about Putnam's Twin Earth see Putnam (1975).

is refreshing, Twin Oscar is entertaining the corresponding belief that twater (XYZ) is refreshing. Both beliefs are true. Oscar and Twin Oscar have corresponding beliefs with the same truth-value and there is a match in the epistemic status of the corresponding beliefs. Oscar and Twin Oscar share their epistemic situation in spite of having beliefs with different content. The lesson learned is that in order to share our epistemic situation zombies do not need to have beliefs with the same content.

According to Chalmers, when I have a belief containing a phenomenal concept that refers to a phenomenal state, the content of the corresponding concept in the zombie's belief refers to some other sort of state, a schmenomenal state.

In that case the materialist can maintain that Chalmers' and Chalmy's corresponding beliefs have the same truth-values and are justified in similar ways. They are just about different things. As Carruthers and Veillet present it:

> Well, on our view zombies are still zombies in that they are not phenomenally conscious. Their perceptual states don't have phenomenal feels. In this respect it is all dark inside. Yet they have something playing a certain role in their psychology – a role analogous to the role that phenomenal consciousness plays in ours. They have something epistemically just as good as consciousness, but they don't have anything that is phenomenally as good. And it seems that this is what matters here. The schmenomenal states they undergo do not feel like anything. (Carruthers and Veillet, 2007)

Chalmers complains that this does not satisfy ES, for it "requires either deflating the phenomenal knowledge of conscious beings, or [...] inflating the corresponding knowledge of zombies" (Chalmers, 2007, p. 20). However, as Carruthers and Veillet argue, it is unclear why one should concede this to Chalmers. When zombieMary sees a rose for the first time, she gains as much knowledge as Mary gains. Mary's knowledge and zombieMary's knowledge are about different things.[42]

The only way I can see to support the claim that Carruthers and Veillet's view would require either deflating Mary's knowledge or inflating

42 Chalmers (2010, fn. 5) maintains, *pace* Carruthers and Veillet, that zombies without such an epistemic situation are conceivable. I see no support for the idea that such zombies are conceivable and I think they are not. Diaz-Leon (2010a) appeals to Hill and Mclaughlin (1999) theory of phenomenal concepts to back up the claim that they are not:

> [P]henomenal concepts and physical concepts play very different psychological roles, and this is what explains the lack of a priori connection. We could characterize these psychological roles in purely functional terms, and therefore a zombie (that is, a functional duplicate of us) would also have concepts that played those different roles. So we could talk about my zombie's corresponding quasiphenomenal concepts (those concepts that are functionally equivalent to my phenomenal concepts) and my zombie's corresponding physical concepts. Since these zombie-concepts also play different roles, they will not be a priori connected, and therefore, sentences involving quasi-phenomenal concepts cannot be a priori inferred from sentences involving only physical concepts.
>
> 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. (ibid. p.13)

ZombieMary knowledge is by maintaining that that the characterization of our epistemic situation entails phenomenal truths a priori.

However, if Q, a phenomenal truth, is part of our epistemic situation, does C has to explain our epistemic situation? Diaz-Leon (2010a) has argued that in this case, C cannot and doesn't have to explain our epistemic situation, materialism is not in danger.

Diaz-Leon argues that what the phenomenal concept strategy has to explain is the a priori disconnection between phenomenal truths and physical truths, not our entire epistemic situation with regard to consciousness, if the epistemic situation includes Q. Chalmers' characterization of the epistemic gap requires phenomenal truths, because in order to be in our epistemic situation with regard to the gap one needs to have phenomenal states and not to be able to infer a priori phenomenal beliefs from a physical description of the world. Diaz-Leon quotes from Chalmers (2007) to stress this point:

> Whereas the inferential disconnection strategy might physically explain an inferential disconnection between physical and phenomenal beliefs, the anti-physicalist's crucial epistemic gap involves a disconnection between physical and phenomenal knowledge. (Chalmers, 2007, p. 24)
>
> In the anti-physicalist's arguments, the relevant epistemic gap (from which an ontological gap is inferred) is characterized in a way that truth and knowledge are essential. [...] It is crucial to the conceivability argument that one can conceive beings that lack phenomenal states that one actually has. And it is crucial to the explanatory gap that one has cognitively significant knowledge of the states that we cannot explain. (Chalmers, 2007, p. 23)

According to this characterization, the truth of the epistemic gap a priori entails phenomenal truths. Diaz-Leon characterizes Chalmers' view on the epistemic gap as: Q & it is not a priori that $(P \rightarrow Q)$. We already know that Q is not explainable in physical terms and the phenomenal concept strategy explains why this is so. Consequently, if the epistemic gap is characterized this way, the type-B materialist will never be able to explain it, because she cannot explain Q in physical terms, but she can explain why this is so.

Chalmers recommends a characterization of the epistemic situation and C that is topic neutral, namely C*, a characterization that is neither C_{phys}, nor C_{phen} because

> This allows the possibility that even if consciousness cannot be physically explained, we might be able to physically explain the key psychological feature and our epistemic situation. (Chalmers, 2007, p. 175)

But this cannot be satisfied if the epistemic gap presupposes phenomenal truths Diaz-Leon (2010a):

(i) P explains C* [it is a priori that $P \rightarrow C*$]

(ii) P does not explain Q [it is not a priori that $P \rightarrow Q$]

(iii) Our epistemic situation E entails Q [it is a priori that $E \rightarrow Q$]

(iv) C* cannot explain E [it is not a priori that $C* \rightarrow E$]

On the one hand, if the characterization of our epistemic situation does not entail Q a priori then it is not clear why zombies do not share our epistemic situation, as we have previously seen. In this case (Anti-PCS) is an invalid argument. If, on the other hand, our epistemic situation entails Q a priori, then the phenomenal concept strategy does not have to explain our whole epistemic situation. As Diaz-Leon claims, what the phenomenal concept strategy has to explain is why if phenomenal properties metaphysically supervene on physical properties there is an inferential disconnection between physical truths and phenomenal truths; i.e. why there is no a priori entailment from physical truths to phenomenal truths.

Materialists should reject Chalmers' characterization of the explanatory gap as Diaz-Leon notes. The explanatory gap should better be characterized as: $Q \rightarrow$ (it is not a priori that $P \rightarrow Q$).[43] The dualist claims that the conceivability of $P\&\neg Q$ jeopardizes materialism, but materialism is only jeopardized if phenomenal consciousness exists. The phenomenal concept strategy can explain the explanatory gap characterized this way. If the phenomenal concept stategy is true, then if there is phenomenal consciousness, then there is an explanatory gap. If the explanatory gap (E) is characterized as $Q \rightarrow$ (it is not a priori that $P \rightarrow Q$): $P\&\neg C$ is not conceivable but C can explain the explanatory gap, because it is a priori that $C \rightarrow E$. That's all the phenomenal concept strategy has to explain.

The first horn

In this horn I will consider the possibility that Q cannot be left out of the epistemic situation. So, zombies do not share our epistemic situation. Furthermore, I will concede that there is a topic neutral characterization of the explanatory gap that does not build phenomenology into the epistemic gap by definition.[44] In this case our epistemic situation should better be cast in phenomenal terms.[45] I will maintain following Balog (mingc) that the truth of (2phen) poses no problem for materialism.

Balog claims that C has to be cast in phenomenal terms and that if the phenomenal concept strategy is true, we should expect precisely C not to be physically explainable.

The microphysical description of the world, P, doesn't explain C_{phen}, but it explains C_{phys}. According to the advocate of the conceptual strategy, C_{phen} and C_{phys} express the very same fact.

---

43 Note the difference between Chalmers' characterization of the gap (Q & it is not a priori that $(P \rightarrow Q)$) and $(Q \rightarrow$ (it is not a priori that $P \rightarrow Q))$.

44 Chalmers (2010) maintains that this can be done in his reply to Diaz-Leon:

> [T]he epistemic gap can be characterized topic neutrally, perhaps along the following lines: we possess a quasi-phenomenal concept q, such that our quasi-phenomenal belief *someone has q* is true [and constitutes knowledge] and *if P, then someone has q* is a priori. (ibid., p.325 fn.4; emphasis in the original)

It is unclear to me how we can make sense of such a proposal, in such a way that there is no explanatory gap for zombies, without building in phenomenology and why such a characterization of the gap would be preferred to the one previously offered.

45 In fact I think that in such a case it can only be cast in phenomenal terms.

To reject the dualist argument, the advocate of the phenomenal concept strategy needs to argue that it is conceivable that both C_{phen} and C_{phys} refer to the very same fact. If it is conceivable then the materialist can deny that one can conclude that C_{phen} is not physical from the lack of entailment, the failure in the explanation, between P and C_{phen}. If there is no a priori reason for ruling out this situation, then materialism has nothing to fear from the gap.

It seems to me that the only way the dualist can block Balog's rejoinder is precisely by assuming that the lack of explanation does a priori entail an ontological gap between the explanans and the explanandum. This would be question begging against the phenomenal concept strategy designed precisely to negate this principle. The dualist cannot merely hold on the principle that maintains that there is a connection between the explanatory gap and an ontological gap to reply to Balog, for this is precisely the principle that the phenomenal concept strategy intends to show to be mistaken.

Chalmers (2007) suggests that this line of reasoning leads to a regress of explanation. The claim that C can be described in phenomomenal and in physical terms merely shifts the problem from the level of phenomenal character to the level of C. The problem that we solved at the level of phenomenal character reappears at the level of C. A new gap would arise at the level of concepts. How can C_{phen} be physically constituted?

This objection is misguided. As Balog points out, the gap at the level of phenomenal concepts is the very same gap that the one at the level of C. The very same phenomenon that explains the fact that phenomenal concepts and physical concepts can be co-referential explains that C_{phys} and C_{phen} are co-referential. This can be nicely illustrated by appealing to the particular theory about phenomenal concepts that Balog has in mind: the constitutive account.

On this account, there is a more intimate relation between a phenomenal concept and the phenomenal character it refers to, more intimate than any causal or tracking relation: the experience itself is constitutive of the phenomenal concept. Phenomenal concepts are constituted by the phenomenal experiences they refer to. If 'having an experience with phenomenal character PC_{RED}' is 'having a neural activity N', then the neural activity N is also constitutive of the concept $\#PC_{RED}$. For most concepts, it doesn't matter what constitutes a particular token of a concept, so long as the right kinds of causal or informational relations hold between it and the rest of the world. For example, it doesn't matter what neural mechanisms constitute a particular token of the concept *fly* as long as the right kinds of causal or informational relations between the particular mechanisms and flies hold. On the other hand, constitution does matter for phenomenal concepts in two senses: in terms of how reference is determined and in terms of how the concepts present their referents. Every token of a phenomenal concept applied to a current phenomenal experience is constituted by a token of the phenomenal experience itself. Different detailed versions of the constitutional account of phenomenal concepts can be seen in Balog (minga); Block (2006); Hill and Mclaughlin (1999); Papineau (2002). Chalmers (2003b) offers also a proposal along these lines.

The constitutive account explains why there is no new gap at the level of C. C_{phen} claims that the *experience itself* is constitutive of the phenomenal concept that refers to it. C_{phys} claims that *the same neural*

activity that constitutes the experience constitutes also the phenomenal concept that refers to it. There is no new explanatory gap at that level. It is the very same that the phenomenal concept strategy intends to explain. As Balog (mingc) notes:

> The ontological implications of the gap between C_{phen} and P have to be denied which comes much to the same thing as denying the ontological implications of the original gap between Q and P. What Chalmers overlooks is that the PCS [Phenomenal Concept Strategy] provides a conceivable physicalist explanation of the conceptual/epistemic gaps (including the new gap involving phenomenal concept descriptions) and so of the intuitive appeal of the anti-physicalist principles – which amounts to more than a mere denial.
>
> In fact, and this is a key point, Chalmers engages in the same kind of circular argumentation against the physicalist that he accuses the physicalist of doing. He rebuts the PCS by assuming that the contested principles are true. So upholding the anti-physicalist principles in the face of physicalist challenges requires an assumption of their truth. (Balog, mingc)

Chalmers (2010, p. 322 fn. 3) complains that the reply explored by Balog in this second horn is a "physicalist explanation" only if we assume as part of it the key claim that phenomenal states are physical states. However, as we have seen, this is the only thing that the phenomenal concept strategy has to do. It is not the job of the phenomenal concept strategy to make the case for materialism.

By endorsing the phenomenal concept strategy, the materialist is only arguing that the anti-materialist arguments do not show that materialism is false, despite what their proponents claim. There is an alternative explanation of the failure of an entailment between physical truths and phenomenal truths. The phenomenal concept strategy is essentially a defensive strategy.

It is not the purpose of this dissertation to show that materialism is true, but to provide a theory of phenomenal consciousness that is compatible with the truth of materialism. In this chapter I have attempted to show that the explanatory gap does not undermine such a project.

2.4 SUMMARY

In this chapter I have presented two classical and related arguments against materialism: the modal argument and the knowledge argument. I have further presented an strategy that replies to both of them: the phenomenal concept strategy.

The modal argument maintains that there is an entailment between a certain form of conceivability and metaphysical possibility. Accepting the very same principles that back up the premises of the argument we can derive the unacceptable conclusion that it is a priori that if some statement about phenomenal consciousness (like I am having a headache now) is true then materialism is false. I have argued, following Balog, that in order to show that this conclusion is false one needs merely to show that one cannot rule out a priori an anti-zombie world.

The remaining work for the materialist is to explain the conceivability of zombies. This work is done by the phenomenal concept strategy.

The knowledge argument similarly exploits the lack of a priori entailment between phenomenal truths and physical truths to show a problem for the explanation of phenomenal truths in physical terms. I have argued that these problems are not exclusive of materialism but common to any reductive theory.

I have discussed three different views on the relation between a priori entailment and reductive explanation. I have concluded that the lack of a priori entailment shows that there is a failure in the explanation exclusive of phenomenal consciousness (or at least not ubiquitous in scientific explanation) and denied that the right conclusion to be derived from this gap is the truth of dualism. To block this conclusion, an explanation of the failure of the a priori entailment has to be provided. This is the task of the phenomenal concept strategy in the last section.

I have finally presented the phenomenal concept strategy. According to the phenomenal concept strategy, the anti-materialist arguments take their force from a misunderstanding on the special nature of phenomenal concepts, the concepts we deploy for referring to the phenomenal character of our experiences. This special nature explains that we cannot see a priori that physical and phenomenal concepts are co-referential. No amount of a priori reflection on phenomenal concepts alone will reveal any physico-phenomenal entailment, even of a contingent type. This fact accounts for the explanatory gap and the conceivability of zombies.

I have presented and rejoined two arguments against the phenomenal concept strategy. The first one, due to Tye, maintains that phenomenal concepts are not special at all. Tye maintains that partial understanding suffices for possession of phenomenal concepts. I have replied that his argument requires that all phenomenal concepts can be deferentially used and I have shown that this is not plausible.

The second argument is due to Chalmers and holds that either phenomenal concepts cannot be explained in a way that is compatible with materialism, or if they can be, then what cannot be explained is our epistemic situation with regard to the gap. I have tried to show, on the one hand, that the only way to deny that our epistemic situation can be explained is by including phenomenal truths in the characterization of our epistemic situation and that a characterization of the explanatory gap that does not presuppose phenomenal truths is to be preferred. If phenomenal truths were part of the characterization of the epistemic situation then materialism would never be able to provide an account of it and the phenomenal concept strategy explains why this is so in spite of the physical nature of phenomenal consciousness. On the other hand, not being able to *reductively* explain in physical terms the psychological capacity for having phenomenal concepts is not a problem for materialism either. The failure in the explanation is an exact reflection of the same phenomenon that the phenomenal concept strategy tries to explain.

Materialism has nothing to fear from these classical anti-materialist arguments if the phenomenal concept strategy is true.

3 PHENOMENAL CONSCIOUSNESS AND VAGUENESS

In the previous chapter I have discussed two classical arguments against materialism: the modal argument and the knowledge argument. These arguments are based on the lack of a priori entailments between physical truths and phenomenal truths. This lack of entailment is exposed by the conceivability of zombies or the intuition that Mary seems to acquire new knowledge when she sees a rose for the first time.

In this chapter, I am going to discuss and reply to a completely different line of reasoning against materialism.

The phenomenal character of the experience is the way it is like for someone to undergo a phenomenally conscious experience. The arguments that I will consider in this chapter maintain either that phenomenal characters are vague and physical properties are not or that phenomenal characters are sharp and physical properties are vague. From this they conclude that phenomenal characters cannot be identified with physical properties. I am going to refer to these arguments as Vagueness based Anti-Materialist (VAM) arguments.

In 3.1, I introduce the phenomenon of vagueness. Vague properties present borderline cases. There are things that neither determinately have the property nor determinately lack it. Proponents of the one form or another of VAM arguments seem to hold a principle that maintains that phenomenal characters and the properties that account for them must present the same borderline cases. In particular, they claim that, if phenomenal characters are vague, the kind of properties that account for them must be vague, and the other way around.

Some philosophers have considered that phenomenal characters are vague whereas physical properties are not. This would jeopardize materialism if their arguments are sound. In section 3.2 I will distinguish two senses in which phenomenal characters can be said to be vague: horizontally and vertically. The former is related to the qualitative character of the experience, the latter to the subjective character. The distinction is relevant because the reasons for claiming that phenomenal characters are vague are completely different in one case and in the other. I will consider some reasons, given in the literature, to believe that phenomenal characters are vague in either sense.

The non-transitivity of the relation 'looks the same as' has been used to support the claim that phenomenal characters are horizontally vague. I will argue that this mistakes the notion of distinguishability that should individuate phenomenal characters (and therefore experiences,[1]) and that it presupposes that cognitive access is essential to the phenomenal character. I will further consider arguments that support the claim that phenomenal characters are vertically vague. I will maintain that these arguments are based either on a confusion on the notion of consciousness in play or on a confusion between metaphysics and epistemology.

1 Recall that the phenomenal character is what makes an experience the kind of experience it is, and a phenomenally conscious experience at all

In the last section, 3.3 I will consider an argument that accepts that phenomenal characters are sharp but not so physical properties. I will argue that this is not a problem for materialism.

3.1 VAGUENESS

We are interested in phenomenal characters. Phenomenal characters are properties of the experience. The vagueness of a property P manifests itself in, at least, one of the following ways:

- There are entities for which it is indeterminate whether they belong or not to the extension of the property P; that is to say, there are borderline cases.
- The Sorites paradox applies to the corresponding predicate: 'X is P'.

An example of vagueness is expressed by the following sentence: 'Sebas is bald'. I like to think about myself as a borderline case of baldness. It is unclear (indeterminate) whether I instantiate the property of *being bald* or not. The property of *being bald* is a vague property. It has borderline cases.

Furthermore, the predicate 'X is bald' is susceptible to a sorites paradox argument. Sorites arguments have the following form:

Base-step Pa_1

Induction-step $\forall n(Pa_n \rightarrow Pa_{n+1})$

Conclusion Pa_m

In these arguments, both premises seem to be acceptable but the conclusion is not. Consider the example of *being bald*:

Base-step A person with no hair is bald

Induction-step If a person is bald, another with one hair more is bald.

Conclusion A person with one billion hairs is bald

The conclusion is clearly false despite the apparent truth of the premises.

There is an important discussion in the literature about the kind of phenomenon that vagueness is; on whether vagueness is a semantic phenomenon, a metaphysical phenomenon, an epistemic phenomenon or a combination of them.

Those who defend that vagueness is a semantic phenomenon (Dummett 1975; Fine 1975; Keefe 2000) maintain that many of our concepts lack a sharp extension. According to the proponents of the semanticist approach to vagueness, for certain entities it is indeterminate whether they fall or not under the extension of the concept. To say that the property P is vague is to say that the predicate 'X is P' has no sharp extension.

The advocates of metaphysical vagueness (Tye 1990) maintain that the ultimate nature of some entities is not sharp. For instance, if what characterizes a certain property, what the property is, is a certain

functional role, then it can be indeterminate whether a certain entity satisfies the functional role.

Last but not least, epistemicism claims that vague predicates have sharp borders but we are cognitively closed to them. There is a precise number of hairs N you can implant to a person that has no hair at all in her head such that, if you add this number of hairs, she is still bald, but if you add N+1 she is definitely not bald. Although vague predicates are admittedly indeterminate in their extension, the indeterminacy is not semantic. This view on vagueness holds that the conundrum presented by the sorites paradox is an epistemological one which in no way undermines classical semantics or logic (Sorensen 2001; Williamson 1996).

In this chapter, I will try to remain neutral on what the ultimate explanation of vagueness is. When this is not possible I will make explicit my departure from this neutrality. In what follows, the claim that a property P is vague should be broadly understood. If, for instance, the semantic view is correct, then the claim that property P is vague should be read as the claim that there is an entity X such that it is indeterminate whether 'X is P' is true.

Understanding the nature of vagueness is a very interesting issue in philosophy, but I am not going to deal with it in this chapter. My aim is to discuss and reply to some arguments against materialism, or particular materialist theories that have appeared in the recent literature, in which vagueness plays an essential role. Those arguments hold either that phenomenal characters are vague and physical properties are not, or that phenomenal characters are sharp and physical properties are vague. They conclude that phenomenal characters cannot be physical properties. The proponents of VAM arguments seem to be endorsing what I am going to call 'the vague identity principle' (V-identity):

V-IDENTITY

> If P and Q are properties:
>
> 'P = Q' is true only if P and Q match in their borderline profiles.

P and Q match in their borderline profiles just in case every borderline case of P is a borderline case of Q and the other way around:

$$\forall x(P(x) \text{ is } \mathtt{borderline} \leftrightarrow Q(x) \text{ is } \mathtt{borderline})$$

V-identity seems to be a plausible principle. However, one could try to reject it by noticing that most of our theoretical identities and paradigmatic cases of theoretical identification seem to fail to satisfy it. Consider the following expressions:

WATER Being water is being H_2O

According to the V-principle, borderline cases of being water are borderline cases of H_2O. The opponent of the principle could reason as follows: "I have bought a bottle of water whose label says: 'This is just water'. As far as I know, no one has complained against the company that produced this bottled water for saying something false, nevertheless the liquid inside is not just H_2O. We want to maintain that WATER is true. So, there are certain levels of impurity in the composition of the liquid that are acceptable for being water. Such a thing is unacceptable in the case of being H_2O. WATER violates the V-identity principle."

This reasoning is completely unappealing. Even in ordinary speech we talk of impurities and foreign bodies of water when the liquid contains something that is not H_2O. Our use of the concept water, as of other natural kind terms, is deferential: we are willing to accept corrections about how to apply the concept by the experts on which we defer our use. If the experts we defer to tell us that being water is being H_2O and that the bottle contains H_2O and an insignificant concentration of calcium we would not strictly accept that the bottle contains just water and consequently the V-principle holds.

We have, I think, good reasons for holding on the V-principle and I will assume that it is true in the rest of the chapter.

3.2 IS PHENOMENAL CONSCIOUSNESS VAGUE?

In this section I will investigate whether the phenomenal characters of our experiences can be said to be vague. I will distinguish two senses in which phenomenal characters can be considered vague and discuss them separately. I will call these two ways in which phenomenal consciousness can be said to be vague horizontal and vertical:

HORIZONTAL-VAGUENESS The phenomenal character Q of an experience of a subject S is horizontally vague if and only if it can be indeterminate whether S's experience has phenomenal character Q or some other phenomenal character.

VERTICAL-VAGUENESS The phenomenal character Q of an experience of a subject S is vertically vague if and only if it can be indeterminate whether S's experience has phenomenal character Q or no phenomenal character at all.

These two forms of vagueness reveal the two senses in which phenomenal characters can be vague. It can be that the qualitative character is vague or it can be that the subjective character is vague. It can be indeterminate whether undergoing the experience is like one thing or other, whether the experience has one qualitative character or other, or it can be indeterminate whether there is something it is like for the subject to undergo the experience at all, whether there is subjective character. The former case entails horizontal vagueness, the latter vertical vagueness.

3.2.1 *Is Qualitative Character vague?*

We have seen that vague properties give rise to predicates that are susceptible of sorites arguments, these arguments are characterized by a failure in the transitivity of the corresponding relational predicate.

Qualitative properties accounts for the difference in phenomenal character of two experiences. If we are interested in the relation between the qualitative character and vagueness, we should investigate whether the relation *same phenomenal character as* is susceptible of a sorites argument, whether *same phenomenal character as* is a non-transitive relation. I will deny that *same phenomenal character as* is non-transitive.

Goodman (1951) was, as far as I know, the first one to consider the relation *looking the same as* to be non-transitive. The relation *looking the same as* is non-transitive if there can be three objects A, B and C such that A looks the same as B and B looks the same as C but A doesn't look the same as C. Wright (1975) showed that indiscriminability has

to be non-transitive if i) phenomenal continua are possible and ii) the human discriminatory powers are finite. A simplified version of the argument is the following:

Let's assume that indiscriminability is transitive. Consider a process of change in respect to some observable property, for example the color, such that there is no seemingly abrupt transition (phenomenal continua). Take any two stages A and B such that A is discriminable from B, and nonetheless close enough to it to warrant that all stages lying in between are either indiscriminable from A or indiscriminable from B (given our limited powers of discrimination); i.e. the intermediate stages will appear to have the same shade of color as A or the same shade of color as B. However, this intermediate stages cannot be indiscriminable from both A and B, because *being indiscriminable from* is supposed to be a transitive relation. The region between A and B will, therefore, be divided into two contiguous subregions, one composed of stages indiscriminable from A, and the other one composed of stages indiscriminable from B. Since A is discriminable from B and indiscriminability is a transitive relation, any stage belonging to the first subregion will be discriminable from any stage belonging to the second subregion. However, in this case a seemingly abrupt change must occur contrary to what we have assumed.

The non-transitivity of perceptual indiscriminability has been used as a basis for arguing for the non-transitivity of the relation *same phenomenal character as.* The idea is that the way things look to me depends exclusively on the phenomenal character of the experience I have while looking at the object. For some philosophers (Byrne (2001); Tye (1997, 2002)) the phenomenal character of the experience cannot vary unless there is a change in the way things look to me.

Fara (2001) argues against the non-transitivity of perceptual indiscriminability, *pace* Wright, showing that there is a tension between the possibility of phenomenal continua and the finiteness of human discrimination power. She claims that there is no reason for accepting the conjunction of the two:

> The only support given for Wright's assumption that our powers of discrimination are (b)-finite[2] was a claim about what it would be natural to assume. Ultimately, I suspend judgment about the (b)-finitude of our discriminatory power, as well as about the existence of phenomenal continua. I would be prepared to accept the truth of either, though I doubt that either question could be decided on the mere basis of inward reflection on the character of our own experience. Still, despite my agnosticism about these claims, my position is that we should deny the conjunction, since first, there is such a straightforward tension between them –(if we really have only finite powers of discrimination, how could there be phenomenal continua?)– and second, when taken together, they have an implausible consequence.

de Clercq and Horsten (2004) object to Fara in defense of Wright's proof. Chuard (2010) accepts Fara's challenge and presents empirical

2 According to Fara, our powers of discrimination are (b)-finite if and only if for some sufficiently slight amount of change (color, sound, position, etc.) we cannot perceive an object as having changed by less than that amount, unless we perceive it as not having changed at all (as having changed by zero amount) (op cit. p. 917).

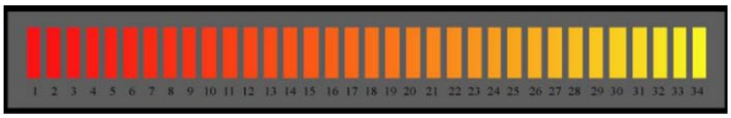

Figure 3: Phenomenal Sorites Chuard (2010).

evidence in favor of the finiteness of our powers of discrimination and phenomenal continua (see figure 3).

My aim in this section is the same as Fara's: to suggest that phenomenal indiscriminability is transitive. But, contrary to her, I will argue that even if their opponents are right in maintaining that perceptual indiscriminability is non-transitive,[3] there are good reasons for resisting the idea that the relation *same phenomenal character as* should be non-transitive.

Perceptual indiscriminability is non-transitive if NTPD is true:

NTPD $\exists a, \exists b, \exists c$, such that [to a subject S with respect to property F]: (a looks the same as b)$\wedge$ (b looks the same as c) $\wedge \neg$(a looks the same as c)

In what follows, I will assume that NTPD is true, and therefore that perceptual indiscriminability is non-transitive. My aim is to show that this does not entail the non-transitivity of the relation *same phenomenal character as*.[4]

Deutsch (2005) appeals to a sorites series to show that phenomenal characters are vague:

> Suppose we divide the spectrum from red to yellow into a series of adjacent patches, and that we divide it finely enough so that, for a normal human subject, each patch looks precisely the same in color as each patch adjacent to it. By so doing, we make it the case that a normal human subject's visual color experience of any particular patch has the same phenomenal character as that subject's visual color experience of any patch adjacent to it. However, the phenomenal character of a subject's experience of the first patch in the series is undeniably different from the phenomenal character of that subject's experience of the last patch in the series. The first patch, which is red, produces an experience with a "red-feeling" character, while the last patch, which is yellow, produces an experience with a "yellow-feeling" character. It follows from what has been said so far that...the relation of same phenomenal character is not transitive. (Deutsch, 2005, pp. 3-4)

From this premise, Deutsch argues that certain materialist theories are wrong, concretely representationalism. To a first rough approximation, we can characterize representationalism as the view that maintains that phenomenal properties are representational properties, and so, the phenomenal character of the experience is determined by its content.[5]

3 I think Fara would disagree with this strategy. Her main motivation is to maintain what she considers a truism: "...if two things look the same then the way they look is the same"(op.cit, p.905)

4 If Fara is right and NTPD is false, the better for me, as it would eliminate this argument in favor of considering phenomenal characters as horizontally vague.

5 I will present representationalism in more detail in the next chapter.

Consider the experience you are having while looking at a certain shade of red, call the phenomenal character of this experience PC_{RED}. According to representationalism, the property of having an experience with phenomenal character PC_{RED} is the property of being in a state that represents a certain property, for example a certain color.

Deutsch maintains that if phenomenal characters are vague and representational properties are not, then phenomenal properties cannot be representational properties. Deutsch endorses the V-identity principle.

I think that Deutsch is misguided in two senses. In this section I will show that we have no reason for accepting that phenomenal properties are horizontally vague, since the relation *same phenomenal character as* is transitive. In the next chapter 4, I will show that for some theories of mental content, intentional properties are vague properties.

I will argue that those who suggest that the relation *same phenomenal character as* is non-transitive misunderstand the notion of distinguishability that should play a role for individuating phenomenal characters. For that purpose let me introduce two different notions of distinguishability. I will call the first one first-sight distinguishability (FS-distinguishability)[6] and third-contrastive (TC-distinguishability) to the second.

FS-DISTINGUISHABILITY Two experiences, E_1 and E_2, are *first-sight distinguishable* (FS-distinguishable or $D_{fs}(E_1, E_2)$) for a subject S if and only if S can distinguish the phenomenal character of E_1 from the phenomenal character of E_2 by simply introspectively comparing the phenomenal character of E_1 and E_2.

It should be clear by now that two things look the same if and only if the experiences I have while looking at them are FS-indistinguishable.

TC-DISTINGUISHABILITY Two experiences, E_1 and E_2, are *third-contrastive distinguishable (TC-* distinguishable or $D_{tc}(E_1, E_2)$*)* if and only if there is an experience *e* such that *e* is FS-distinguishable from E_1 and not FS-distinguishable from E_2.
$D_{tc}(E_1, E_2) \leftrightarrow \exists e(D_{fs}(E_1, e) \wedge \neg D_{fs}(E_2, e))$[7]

If NTPD is true then FS-indistinguishability ($\neg D_{fs}$) is a non-transitive relation: E_1 and E_2 can be FS-indistinguishable, as can E_2 and E_3 without thereby E_1 and E_3 being FS-indistinguishable. On the other hand, TC-indistinguishability is transitive. $\neg D_{tc}$ is a transitive relation, if an experience E_1 is TC-indistinguishable from an experience E_2 and E_2 is TC-indistinguishable from an experience E_3, then E_1 and E_3 are TC-indistinguishable:
$\forall E_1, E_2, E_3((\neg D_{tc}(E_1, E_2) \wedge \neg D_{tc}(E_2, E_3)) \rightarrow \neg D_{tc}(E_1, E_3))$:

(1) $\neg D_{tc}(E_1, E_2) \wedge \neg D_{tc}(E_2, E_3)$ Assumption

(2) $D_{tc}(E_1, E_3)$ Assumption

(3) $\forall e(D_{fs}(E_1, e) \rightarrow D_{fs}(E_2, e))$ From 1 and TC-distinguishability[8]

(4) $\forall e(D_{fs}(E_2, e) \rightarrow D_{fs}(E_3, e))$ From 1 and TC-distinguishability

6 It is important to note that by first sight distinguishability I do not mean prima facie distinguishability. It may well be that distinguishing two experiences at first sight requires plenty of concentration and attention.

7 Note that if NTPD is false, then FS-distinguishability and TC-distinguishability are coextensive.

8 Note that both D_{tc} and D_{fs} are commutable: $D(a, b) \leftrightarrow D(b, a)$

(5) $\exists e(D_{fs}(E_1, e) \wedge \neg D_{fs}(E_3, e))$ From 2 and TC-distinguishability

(6) $D_{fs}(E_1, p) \wedge \neg D_{fs}(E_3, p)$ From 5

(7) $D_{fs}(E_2, p)$ From 6 and 3

(8) $D_{fs}(E_3, p)$ From 7 and 4

(9) $\neg D_{tc}(E_1, E_3)$ From 8, 6 and 2 by *reductio ab absurdum*

$\therefore$ $(\neg D_{tc}(E_1, E_2) \wedge \neg D_{tc}(E_2, E_3)) \rightarrow \neg D_{tc}(E_1, E_3)$ From 1-9 by $\rightarrow$introduction

I have presented two notions of indistinguishability: FS-indistinguishability and TC-indistinguishability. The question that I will address in what follows is which notion of distinguishability we should prefer for the individuation of phenomenal characters. If it is the notion of FS-distinguistability then the relation *same phenomenal character as* will be non-transitive. If, on the other hand, it is the notion of TC-indistinguishability then the relation *same phenomenal character as* will be transitive.

Let E_1 and E_2 be two numerically different experiences and PC_1 and PC_2 their respective phenomenal character.

FS-INDIVIDUATION Two phenomenal characters are the same if and only if the corresponding experiences are FS-indistinguishable: $\forall n \forall m(\neg D_{fs}(E_n, E_m) \leftrightarrow PC_n = PC_m)$

TC-INDIVIDUATION Two phenomenal characters are the same if and only if the corresponding experiences are TC-indistinguishable: $\forall n \forall m(\neg D_{tc}(E_n, E_m) \leftrightarrow PC_n = PC_m)$[9]

Deutsch commits himself to FS-individuation. He maintains that FS-indistinguishability does suffice for experiences to have the same phenomenal character.[10]

On the other hand, philosophers like Goodman (1951) would be willing to accept TC-individuation:

> [T]he visual experiences of two objects have the same phenomenal character just in case they look the same as each other and look the same as all the same third parties as well.

Deutch's motivation for defending FS-individuation is that if TC-individuation is true, one cannot tell just by introspecting whether two experiences share a phenomenal character. He maintains that there is a conceptual connection between visual indistinguishability and sameness of visual phenomenal character.

The view that FS-distinguishability should be used as the individuation criterion for phenomenal characters is often held and intuitively appealing. It is intuitively appealing that the way the things looks to me depend only on the phenomenal character of the experience I have while looking at these things. The intuition is, as I will try to show, nonetheless wrong: the way things look depends also on the access we have to the phenomenal character of our experiences. If the access we

9 Note that if two experiences are FS-distinguishable, then they are TC-distinguishable and do not share the same phenomenal character.

10 Dummett (1975) seems to agree with him.

have to the phenomenal character of our experiences is less fine-grained than the phenomenal character itself, then two objects can look the same to me whereas the experiences I have while looking at them differ in phenomenal character. This idea requires further clarification.

In the first place, I should argue in favor of the claim that the way things look does not depend exclusively on the phenomenal character of the experience: two experiences may have different phenomenal characters and nonetheless that the way the world looks to me when I have these experiences is the same.

According to NTPD, there can be two experiences that are not distinguishable at first-sight but are third-contrastive distinguishable. Consider three numerically different experiences, E_1, E_2 and E_3 that satisfy NTPD; neither E_1 and E_2 nor E_2 and E_3 are FS-distinguishable. However, E_1 and E_3 are FS-distinguishable and consequently E_1 and E_2 are TC-distinguishable.

We individuate type-experiences by the way it *feels* to undergo the experience, namely, by their phenomenal character. When I look at my red apple I have an experience E_1 with a phenomenal character PC_{RED}. I blink my eyes and have another numerically different experience E_2. Assume that undergoing E_2 feels exactly the same way undergoing E_1 feels. E_1 and E_2 have the same phenomenal character PC_{RED}. We say that E_1 and E_2 are two different tokens of the same type of experience. In general two experiences are tokens of the same type if they have the same phenomenal character, if the way it feels to undergo them is the same.

In order to decide whether E_1 and E_2 are TC-distinguishable we are exclusively appealing to the way it feels to undergo these experiences, namely to the phenomenal character of the experience. TC-distinguishability relies on phenomenology (on the phenomenal character) alone to differentiate two experiences. If the fact that E_1 and E_2 are not FS-distinguishable makes them have the same phenomenal character, then how can it be that these experiences are TC-distinguishable? If E_1 and E_2 have the same phenomenal character how is it that I can phenomenologically distinguish E_1 from E_3 but not E_2 from E_3?

The proposal I make here agrees with Goodman and denies what Fara and Deutsch consider to be a truism; namely, that if two things look the same to a subject S, then the phenomenal character of the experiences (if the experiences are veridical) S has while looking at these objects is the same. The claim that this is a truism is based on an appealing intuition. The intuition is based on the idea that we have a perfect access to the phenomenal character of the experience. If I cannot tell two experiences apart then they have the same phenomenal character. However, the judgments I make to compare the phenomenal character of two experiences require access consciousness, cognitive access to the phenomenal character of our experiences. If access consciousness and phenomenal consciousness are two different properties then the intuition is unsupported. In 1.2.1 we have seen that there are good reasons to believe that they are different properties. In 5.2 I will further argue that there is empirical evidence that suggests that we can have phenomenally conscious experiences without cognitive access.

Now we have the resources to explain what happens in the case of experiences that satisfy NTPD. If phenomenal consciousness can be dissociated from the cognitive access then it is plausible than the access we have to the phenomenal character of the experience is less fine

grained that the phenomenal character itself. I will not be able to decide, in certain cases, whether two experiences have the same phenomenal character or not. My resources for doing so are limited; the access I have to the phenomenal character of my experience does not allow me to tell whether E_1 and E_2 have the same phenomenal character. If I am right, they do not have the same phenomenal character and that explains that I can tell E_1 and E_3 apart. My access resources to the phenomenal character of these experiences do allow me to distinguish these two phenomenal characters that are sufficiently different. How things look to me depends on the access I have to my phenomenal character and that is why the object I look at when undergoing E_1 looks the same as the object I look at when undergoing E_2, in spite of the fact that the phenomenal character of E_1 and E_2 are different.

My opponent intended to show that phenomenal characters are vague due to the failure of transitivity in the relation *looks the same as*. However, as I have argued, the acceptance of this failure of transitivity does not commit oneself to the claim that phenomenal characters are vague.

In what follows, I will further argue that my proposal is to be preferred to the one of my opponent for two reasons: no further properties have to be postulated and no paradoxical result follows from my proposal.

On the other hand, my opponent could insist that it is preferable to save the intuition that things cannot *look the same* to me unless the experiences I have while looking at them have the same phenomenal character. For that purpose, we should hold that the access I have to my phenomenal character is as fine grained as the phenomenal character itself and consequently phenomenal properties are vague. I completely disagree and I see no reason for holding such intuition once we perfectly explain it away as I have shown. Be that as it may, my opponent owes us an explanation of how can we TC-distinguish E_1 and E_2. She has to postulate an additional property E_1 has and E_2 lacks and an ability we have to access this additional property in virtue of which we can tell apart E_1 from E_2.

In my view, we have a criterion for telling E_1 and E_2 apart: the way it feels to undergo E_2 is, whereas the way it feels to undergo E_1 is not, similar enough (so that we cannot FS-distinguish them) to the way it feels to undergo E_3. The way it feels to undergo E_1 is different from the way it feels to undergo E_2. E_1 and E_2 have different phenomenal characters.

What is more, vague properties have well known problems: they lead to paradoxical results. In this case, if phenomenal characters are individuated by the FS-indistinguishability of the experience, then two experiences that are FS-indistinguishable but not TC-indistinguishable share and do not share a property, the phenomenal character, which is a contradiction. Formally:

P1 $D_{tc}(E_1, E_2)$
P2 $\neg D_{fs}(E_1, E_2)$

(1) $\forall n \forall m(\neg D_{fs}(E_n, E_m) \leftrightarrow PC_n = PC_m)$ Assumption

(2) $\exists E_3(D_{fs}(E_3, E_2) \wedge \neg D_{fs}(E_3, E_1))$ From P1 and TC-distinguishability

(3) $D_{fs}(E_3, E_2)$ From 2

(4) $PC_3 \neq PC_2$ From 1 and 3 by modus tollens

(5) $\neg D_{fs}(E_3, E_1)$ From 2

(6) $PC_3 = PC_1$ From 5 and 1

(7) $PC_1 \neq PC_2$ From 6 and 4 and Leibniz's law.

(8) $PC_1 = PC_2$ From P2 and 1

$\therefore$ $\neg\forall n\forall m(\neg D_{fs}(E_n, E_m) \leftrightarrow (PC_n = PC_m))$ From 1, 7 and 8 by *reductio ad absurdum*

Modern theories of vagueness try to block sorites-like arguments with minimal restrictions on classical logic, like epistemicism or supervaluationism.

For there to be a link from the non-transitivity of the relation *looks the same as* to the claim that the relation *same phenomenal character* is non-transitive, phenomenal characters should be individuated by FS-distinguishability (the assumption of the argument).

One way of resisting the argument would be by denying the validity of Leibniz's Law and particularly the less controversial conditional of the law,[11] the indiscernibility of identicals. This principle holds that if two entities are identical then they share all their properties: $\forall x\forall y[x = y \rightarrow \forall P(Px \leftrightarrow Py)]$[12]

My opponent could block this argument by rejecting the indescernibility of identicals. I consider that, if the proponent of FS-discrimination as a criterion for the individuation of phenomenal characters is committed to the rejection of this principle, we have good reasons for rejecting his view, especially if we have an alternative. Appealing to TC-indistinguishability as an individuation criterion for phenomenal character is such an alternative. It explains that two experiences are not distinguishable by introspection without thereby entailing that they have the same phenomenal character. The reader can weigh the cost of giving up Leibniz's law versus giving up simple introspection as a criterion for phenomenal character individuation.

The fact that we cannot FS-distinguish between two experiences does not show that both have the same phenomenal character, as we have seen. I have maintained that the failure in transitivity of the relation *looks the same as* is not a good reason for believing that s*ame phenomenal character as* is not transitive.

My opponent has to postulate an additional property to explain TC-distinguishability. On his account, phenomenal characters do not suffice for TC-distinguishability and we can tell apart two experiences that are not distinguishable at first sight. It is obscure what those properties would be if they are not the phenomenal character of the experience. On the other hand, the relation 'look the same as' is non-transitive

11 The Leibniz law is an ontological principle that holds that entities are identical if and only if they share all their properties. It is composed by two conditionals: the identity of indiscernibles (if two entities are indiscernible then they are identical) and indiscernibility of identicals (if two entities are identical then they are indiscernible). In order to show that the identity of indiscernibles is false, it is sufficient that one provides a model in which there are two non-identical entities having all the same properties. Max Black (Kim and Sosa (1999)) claimed that in a symmetric universe wherein only two symmetrical spheres exist, the two spheres are two distinct objects, even though they have all the properties in common. On the other hand, the indiscernibility of identicals is usually taken to be an uncontroversial claim.

12 In our case we are interested in properties, so x and y are properties and $\forall P$ quantifies over properties of properties.

(if NTPD is true) and my opponent proposal secures the view that how things look depends on phenomenal characters. But this view is completely compatible with my proposal that TC-distinguishability individuates phenomenal characters. How things look depends also on the phenomenal character of the experience I have when I look at the object but not exclusively on it. How things look like depends on the cognitive access we have to the phenomenal character of our experience. This access does not allow us to distinguish two experiences with very similar phenomenal character.

In this section I have tried to show that we have no reason for thinking that phenomenal characters are horizontally vague. In the next section I will hold that we have no reason for believing that phenomenal properties are vertically vague either.

3.2.2 *Is Subjective Character vague?*

The question as to whether the subjective character is vague is directly related to the question as to whether phenomenal character can be vertically vague. The subjective character is vague if and only if there are borderline cases of phenomenally conscious experiences. I am going to argue that we have no reason for believing that phenomenal characters are vertically vague.

It seems to me that phenomenal consciousness does not admit borderline cases. Intuitively phenomenal consciousness is not vague; for a given feeling, A, an experience with a phenomenal character A, you either definitely have this feeling or you definitely don't. Phenomenal consciousness is a matter of on/off. A mental state of a subject S is either definitely phenomenally conscious or it is not; either it contributes to what it is like for S to undergo certain experience or it doesn't.

Michael Antony (2006a) has given arguments that go beyond the mere intuition to maintain that phenomenal consciousness cannot be vague. I do not find them very compelling.

Antony argues that vague predicates are susceptible to sorites series and the elements in a sorites series have the following features:

1. There is a feature F such that the elements in the sorites series vary in F.

2. F is closely tied to the notion we have of the vague term.

For example, in the case of baldness, subjects in the sorites vary in something like the quantity of hair or the quantity of hair with regard to the head surface, and this conception is tied to our notion of 'bald'. Antony argues that, in the case of phenomenal consciousness, there is no such a feature F such that: (i) elements in a sorites series vary in F and (ii) F is closely tied to our notion of phenomenal consciousness.

I find Antony's argument unappealing because premises (1) and (2) seem to be false. There are many cases of vague predicates such that there is no such a feature F that satisfies both 1 and 2. Consider the case of *being intelligent*. *Being intelligent* is a vague property but nevertheless there seems not to be a feature F such that we can build up a sorites series varying it.

If the reader finds Antony's argument compelling, then the better for my purposes. I will simply rest on the intuition that phenomenal consciousness is sharp. The burden of the proof is on my opponent

if he wants to make an argument against materialism based on the vague nature of phenomenal consciousness. Some philosophers have tried to provide such an argument. I will consider some reasons for considering that subjective character is vague and argue that those reasons are unsound.

David Papineau (2002) has defended that phenomenal consciousness is vague. He conceded, in previous work, the intuition that phenomenal consciousness admits no borderline cases:

> When we look into ourselves we seem to find a clear line. Pains, tickles, visual experiences and so on are conscious, while the processes which allow us to attach names to faces, or to resolve random dot stereograms are not. True, there are "half-conscious" experiences, such as the first moments of waking, or driving a familiar route without thinking about it. But, on reflection, even these special experiences seem to qualify unequivocally as conscious, in the sense that they are like something, rather than nothing. Papineau (1993, p. 125)

Papineau, nevertheless, argues that the intuition is wrong. He claims that the idea of phenomenal consciousness being sharp, that feelings are either present or they are not, is due to the dualist idea that phenomenal consciousness is some kind of inner light.

We can think of experiences that are very *vivid*, like a horrible headache or the smell of the coffee just brewed. If you take a pain killer and wait or you stay at a prudent distance of the coffee for hours, the intensity of the experience vanishes. Independently on how impoverished the phenomenal character is, as far as there is any experience, there is phenomenal consciousness. It might be indeterminate whether the experience has one phenomenal character or other (horizontal vagueness), but for any mental state of a subject S either determinately there is something it is like for S to be in that state or determinately there isn't. The inner light intensity can reduce, but for any level the light is fully extinguished or it isn't. This is, according to Papineau, what supports the intuition that phenomenal characters are not vertically vague.

> If you accept this dualist intuition, then you will think that it must be determinate whether phenomenal consciousness is present or not. If consciousness is an extra inner light, so to speak, distinct from any material properties, then there must always be a definite fact of the matter whether this light is switched on, however dimly, even in unfamiliar cases. Papineau (2002, p. 203)

Papineau seems to admit that in the case of humans, phenomenal consciousness is sharp. But for other beings that are unable to think about their mental states (like sharks or octopuses) there will be no way of deciding which states are phenomenally conscious. Papineau concludes that "we should accept that sometimes it will be a vague matter which states of which beings are conscious." (ibid. p.125)

Papineau seems to suggest that there could be no fact of the matter on whether beings that are physically different from us have phenomenally conscious experiences.

> It may seem very odd to hold that a phenomenal term like 'seeing something red' is vague, and that there is therefore

> no fact of the matter of whether a silicon doppelganger looking at a ripe tomato is seeing something red or not....My claim is not that it is vague how it is for the doppelganger. The doppelganger's end experience will feel as it does, and there is no need to suppose that this in itself is less than definite, that there is somehow some fuzziness in the doppelganger's experience itself. Rather, my claim is that our phenomenal term 'seeing something red'...is not well focused enough for it to be determinate whether or not the doppelganger's experience falls under it ... when we seek to apply the term beyond the cases where it normally works, it issues no definite answer... There is no reason to suppose that there is anything in the workings of the term to decide this question. (Papineau, 2002, pp. 199-200)

If we are realists about phenomenal consciousness, then what phenomenal properties are, the metaphysics of phenomenal properties, is independent of our knowledge of them. As a realist we should distinguish metaphysics and epistemology. Consider the following claim:

(DATA) Commander Data is having a sensation as of red

The proposition expressed by (DATA) is determinately true or determinately false. It can be the case that we can by no means know its truth value, but as realists, it does have determinately one. If there is no a priori conection between phenomenal truths and physical truths, no analysis of our phenomenal concept of sensation of red (PC_{RED}) can help us to decide whether Data is having sensation of red or not. But that doesn't show that phenomenal consciousness is vague. Our phenomenal concept *sensation of red* (PC_{RED}) is such that it allow us to conclude that either Data is determinately having sensation of red or he is determinately not having *sensation of red*. If someone would like to conclude that sensation of red is vague from the fact that we cannot come to know whether a being is having the sensation of red or not he would be making a mistake.

All that Papineau's example shows is an epistemic problem, a problem already voiced by Block (2002a), the harder problem (see 2.1.2): we cannot know whether there is something it is like for Commander Data to see my red apple. This point can be acknowledged by the materialist without any appeal to the vagueness of our notion of phenomenal consciousness. Papineau's intention is in fact to offer his analysis as a solution to the harder problem:

> I agree with Block that this indecision [as to whether non-human creatures are conscious] is a consequence of the inflationist [i.e., phenomenal realist] recognition of phenomenal concepts. However, I don't agree that this represents some kind of deficiency in inflationist materialism. In my view, it is indeed not always possible to answer such question as whether... robots... can feel phenomenal pain....One possibility [why that is so] is that questions about phenomenal consciousness always have definite answers, but epistemological obstacles bar our access to them....But... another possibility... is that our phenomenal concepts are vague. I shall be arguing for this analysis. (Papineau, 2002, p. 178)

We have no reason for believing that Commander Data's experience is a borderline case of phenomenal consciousness. Papineau requires an independent reason to claim that phenomenal characters are vague. He acknowledges that:

> [T]he intuitively more natural view is surely that either doppelgangers or duplicates will have the relevant experiences, or they won't. In the absence of independent arguments for vagueness, it would seem that Block is justified in his claim that inflationists have saddled themselves with an inexplicable barrier to discovery (Papineau, 2002, 198).

Papineau faces the challenge and claims that the correct theory of phenomenal concepts will deliver the result that phenomenal concepts are vague (Papineau maintains that the nature of the vagueness of phenomenal characters is semantic). Papineau is a materialist who embraces the phenomenal concept strategy (see 2.3). He argues that there is no fact of the matter about the level of abstractness at which we should look for the material referent of phenomenal concepts. According to Papineau, the harder problem is not an epistemic problem but a semantic problem. The problem of this proposal, however, is that it would prevent any mind-body identity. Bermudez (2004) presents the objection as follows:

> On the one hand we are told, on the basis of an argument from the completeness of physics, that we can be sure that every phenomenal property is identical to some material property, and therefore that every phenomenal concept refers to some material property. On the other hand, however, we are told that there is no fact of the matter about which material property that might be, for any given phenomenal concept. So, in virtue of the first claim we are told that for any given phenomenal concept P there must be a true identity claim involving it of the form P = M where M is a material concept identifying a material property. At the same time, however, the vagueness of phenomenal concepts (as Papineau interprets it) entails that there is no fact of the matter determining the truth or falsity of any claim of the form P = M. How can the general identity thesis be true when there is no fact of the matter as to the truth of any particular identity claim? (ibid. p.136)

If Bermudez is right, we have good reasons for rejecting Papineau's view on the harder problem.[13] The intuition further supports the epistemological reading of the harder problem. We have the intuition that Data determinately has or determinately lacks phenomenal consciousness, an intuition that it is hard to give up. I take that intuition to be a good reason for preferring a theory of phenomenal concepts that does not commit us to the conclusion that phenomenal characters are vertically vague. Tye (1996) makes a similar point in favor of the intuition:

> This seems to me a pretty amazing view. Maybe Papineau is a color madman. If so, he'll have to concede that it's

13 See also Antony (2006b) for a rejection of Papineau's argument. According to Antony's interpretation of Papineau's theory of phenomenal concepts, if phenomenal concepts seem sharp then they are sharp. Antony maintains that they seem sharp and therefore that they are sharp.

> wholly arbitrary to say that what it is like for him, as he holds up a ripe tomato, is not (or is) the same as for me in the same circumstances. But surely either the phenomenal quality that is present in his experience is present in mine or it isn't. How could there be any arbitrariness here? If it is determinate what his experience is like (and that he knows from introspection) and it is determinate what my experience is like (and that I know in the same way), how could there be any real indeterminacy as to whether his experience is phenomenally distinct from mine? (ibid. p. 685)

I think that Papineau fails to make a compelling case in favor of vertical vagueness.

A completely different argument to the same effect has been presented by Brogaard. She claims that 'conscious' is a relative gradable adjective and she claims that relative gradable adjectives typically are associated with an implicit or explicit standard of comparison that gives rise to borderline cases and triggers the Sorites series.

> [R]elative gradable adjectives...give rise to borderline cases. In the neutral sense of 'borderline case', a borderline case is an individual which does not evidently fall under the predicate and which does not evidently not fall under the predicate. For example, a 20 m2 apartment is clearly tiny even for New York standards, whereas an 800 m2 apartment clearly is not tiny but it may be indeterminate either epistemically or semantically whether a 45 m2 apartment is tiny for New York standards.
>
> [R]elative gradable adjectives...give rise to the Sorites Paradox in their unmarked form, for instance:
>
> > 1. An 800 m^2 apartment is a huge apartment for New York standards
> >
> > 2. If an apartment that is n m2 is a huge apartment for New York standards, then an apartment that is n-1 m^2 is a huge apartment for New York standards
> >
> > 3. An apartment that is 0 m^2 is a huge apartment for New York standards

It is not my aim to discuss the relation between relative gradable adjectives and vagueness. I think that Brogaard can be shown wrong without getting into this discussion.

In the first chapter, I presented a distinction between different senses in which the predicate 'X is conscious' is used in common language. I was careful to make the distinction to focus on the interesting one. I called it phenomenal consciousness. I think that Brogaard overlooks that distinction. Consider some of the examples given by her:

(a) Experts say that up to 40 percent of those thought to be in a persistent vegetative state are, in fact, quite conscious [for someone thought to be in a persistent vegetative state]

(b) Although clinically vegetative and still unable to communicate or respond in any way, the British woman is quite

conscious [for someone who is clinically vegetative and unable to communicate or respond in any way]

(c) Freud came to view dream activity as highly conscious [for brain activity taking place during sleep]

(d) Mary is highly conscious

(e) Mary is highly conscious for someone in a meditative state

(f) Mary is highly conscious for someone concentrating hard on a logic exercise

It is unclear which sense of the predicate is involved in either case, but it is clear that it not the same across examples. Is it the sentience sense, the awaken sense, the phenomenal sense? In the sense of awaken or sentience, I already made clear in 1.2.1, when I introduced the notion of creature consciousness, that sentience admits plenty of borderline cases, but if we consider the phenomenal sense the last three statements make no sense.

I think that Brogaard, and maybe the writer of the article from which she quotes, just misunderstand the notion of consciousness in play. In the case of vegetative states, what is at issue is either the level of sentience or the possibility of having some phenomenally conscious mental states. There is a worry, raised in the last years on whether patients that were thought to be in coma or vegetative state had phenomenally conscious experiences.[14] But the only senses in which the use of consciousness can be considered vague are with respect to the number of patient's conscious states or the content of those states (the amount of information those states track).

Furthermore, all the examples, but (c), are examples of creature consciousness. Even if one could argue that there are examples where the term 'conscious' is used in the phenomenal sense as a relatively gradable adjective, we are talking about a property of the creature. It seems natural, if not obvious, to maintain that any vagueness in this respect comes from the number of phenomenal states the creature instantiates. It could perfectly be the case that in order to qualify as a phenomenally conscious creature, instantiating a single phenomenally conscious mental state does not suffice. It could perfectly be that it is indeterminate how many phenomenally conscious mental states are required for being a conscious creature. What would be relevant for materialism is whether there are borderline cases of phenomenally conscious mental states. That is not shown at all by Brogaard's examples.

Alternatively, the case in favor of the existence of vertically vague experiences can be made by appeal to *dull* experiences. Tye (1996) considers a sense in which the subjective character can be said to be vague.

> But can it be vague whether a given state is an experience, whether there is anything at all it is like to undergo the state? It seems to me that it is not pre-theoretically obvious that the answer to this question is 'No'. Suppose you are participating in a psychological experiment and you are listening to quieter and quieter sounds through some headphones. As the process continues, there may come a point

14 See Laureys and Tononi (2008) for a recent review of neuroimage studies on these patients.

> at which you are unsure whether you hear anything at all. Now it could be that there is still a fact of the matter here (as on the'dimming light model); but equally it could be that it is objectively indeterminate as to whether you still hear anything. So, it could be that there is no fact of the matter about whether there is anything it is like for you to be in the state you are in at that time. In short, it could be that you are undergoing a borderline experience. (ibid. pp. 682-683)

That seems to be a prima facie candidate for being a borderline case of phenomenal consciousness. My intuition is nevertheless that this is not a borderline case. In every instance either there is something it is like for you to hear the tone or there isn't. The case for my opponent rests, I think, on something like the following argument:

(1) The phenomenal character of S's experience is a matter of how things seem to S.

(2) There are S's experiences such that S cannot determine whether it seems somehow to her or not.

(3) If S cannot determine whether it seems somehow to her or not then it is indeterminate whether it seems somehow to her or not.

∴ There are experiences such that it is indeterminate whether they have phenomenal character (whether they are phenomenally conscious).

I think that the argument is unsound. I can see two reasons one could have to accept premise (3). The first one is similar to the one we have previously seen: a confusion between metaphysics and epistemology. The fact that S cannot decide whether being in M seems somehow to him does not, per se, entail that it is indeterminate. The second one is the claim that cognitive access is essential to phenomenal consciousness. This last claim is controversial, and I think wrong. As I have argued, unless one endorses it, premise (3) is left unsupported.

In this section I have discussed some reasons for considering that phenomenal characters are vague in either sense and I have rejected them. In the next section I will consider the opposite: arguments that maintain that phenomenal characters are sharp whereas any plausible candidate the materialist can appeal to is vague.

3.3 PHENOMENO-PHYSICAL IDENTITIES AND VAGUENESS

If the V-identity principle is true, then the phenomenal property and the physical candidate to be identifies with it have to satisfy the same borderline profile if they are going to be identified. Based on this idea some philosophers (Antony (2006a); Deutsch (2005)) have argued against a certain kind of identification or reduction. In its more general formulation the argument against the reduction has this form:

(1) Phenomenal property Q has borderline profile q

(2) Physical property P has borderline profile p

(3) p≠q

(C) P≠Q From 3 by V-identity

In general VAM arguments maintain that one of the two properties involved in the reduction is vague whereas the other is sharp. Trivially that entails a difference in the borderline profiles and if the V-identification principle is true, then the reduction is false.

As we have seen in section 3.2.1, some philosophers have argued against the reduction of phenomenal properties to other properties like representational properties based on the V-identity principle. I have argued there that the representationalist can reject the premise that qualitative character is vague, so the position is not jeopardized.

In section 3.2.2 I have presented some arguments from the literature for sustaining that the subjective character is vague, that having a phenomenally conscious experience is not a matter of on/off. I have shown that there is no reason for maintaining that the phenomenal character is not sharp.

In this section, I will deal with those philosophers who argue that phenomenal consciousness is sharp but any candidate the materialist has available for identification is vague. Given this premise and the V-identification principle, materialism is false, or so argues my opponent. In my defense of materialism I will accept the V-identity principle and deny the premise that any candidate for identification is vague.

Michael Antony argues that phenomenal consciousness is sharp. I think that the argument is unsound, as I said above. However, it seems to me to be the most plausible option and I will grant this premise: phenomenal consciousness is sharp. In Antony (2006a), he maintains that any plausible candidate the materialist has available for reduction is vague. So, according to the V-principle any identification between a phenomenal property and a physical property is false. I will try to show that Antony's argument is unsound.

Antony first considers the case of identification between phenomenal properties and neurophysiological states. For instance the identification of pain with certain pyramidal activity or of a red after-image with activity N involving V1 and V4, etc. What makes a phenomenally conscious state a phenomenally conscious state at all, the subjective character, will be a certain property N.

> Identity theorists focus on determinate types of conscious states, C1, C2 , . . ., Cn (pain, orange after-image, etc.), and neurophysiological states, N1, N2, . . ., Nn (c-fiber firings, activity in visual area V1, etc.), and claim that C1=N1, C2=N2, . . ., Cn=Nn. They typically do not explain what it is in general to be in a conscious state (call that property 'C', which is short for 'conscious state'). However, identity theorists must believe there is some neurophysiological story to be told about what distinguishes conscious from non-conscious states. That story, in effect, will ascribe a single property N to all conscious states, which the identity theorist will identify with C. There are various possibilities for what N might be: a property common to each of N1, N2, . . ., Nn; a disjunction of N1, N2, . . ., Nn; a disjunction of properties more general than N1, N2, . . ., Nn but less general than

> N; and so forth. For our purposes it does not matter how exactly the story goes, so long as some N is identified with C.Antony (2006a, p. 521)

Antony requires a further assumption for his argument. He requires N to be vague. According to the materialist, N is the property we identify with Q. Antony argues that, given the complexity of neuroscience, any candidate will admit borderline cases. For illustration, Antony considers the example of our concept NEURON.

> Neurons are highly complex structures, with diverse components that perform sophisticated micro-functions. Anyone minimally familiar with such details can convince oneself that by gradually removing atoms (or other sufficiently small parts) from such neuronal components, one will eventually reach borderline cases for concepts of many of those components (and their properties), and as a result borderline cases for neuron as well: structures that are neither clearly neurons nor clearly not neurons. (ibid. p.522)

If neurons admit borderline cases and N is made out of neurons, arguably, N will admit borderline cases. We just have to replace determinate cases of neurons for borderline cases of neurons to get a borderline case of N. If N is vague and phenomenal consciousness is not, then the identification is false, according to the V-principle.

I see two possible replies to Antony's argument.

In the first place, the materialist can complain that Antony wants to derive metaphysical consequences from a neutral understanding of vagueness. Our concept NEURON fails to get a sharp reference. There will be objects for which it is indeterminate whether the concept NEURON applies to them or not. Our concept NEURON evolves[15] as we make new discoveries; the concept's precisification increases, and the number of borderline cases diminishes. The problem of this line of reply is that phenomenal consciousness is sharp so the only level at which the concept NEURON would not admit any borderline case is the level of fundamental microphysics, strings if string theory is true. In that case, the identification between phenomenal consciousness and N is an identification at the level of microphysics. Antony is happy with this conclusion, but as he notes, many materialists should not be, for the kind of materialism that remains untouched is too close to the familiar kind of protopanpsychism, neutral monism.

> [T]he appropriate response to the above arguments is to investigate versions of the identity theory or dualism that appeal to physical properties whose concepts are sharp ... In seeking physical properties whose concepts are sharp, the obvious place to look is fundamental physics. Notice, however, that if the nature of consciousness resides at that level, the likelihood that panpsychism is true would appear to increase dramatically. (ibid. p. 531)

If we have to appeal to fundamental physics for finding a candidate to be identified with phenomenal consciousness, then it seems very

15 I use 'evolve' to indicate that the concept is the same and what we do as our science improves is to precisify its reference. One could claim that this change will entail a change in reference and therefore a completely new concept. In that case the identification between phenomenal consciousness and N would be false and this reply is not acceptable.

plausible that everything is made out of the same fundamental particles which would have proto-phenomenal properties (see Chalmers (2003a)).

There is an alternative reply that is much more interesting: functionalism. A neuron is the entity that satisfies a certain functional role.[16] If the functional role is sharp, then there is no problem for the identification with phenomenal consciousness. It may be that the functional role that determines what is a neuron is not perfectly precisified in such a way that our concept of neuron is not vague, but there is no reason for believing that it won't be.[17]

Antony anticipates this reply and argues also against functionalist views in general. For functionalists, phenomenal consciousness is identified with a certain functional role. There is a certain functional role that all phenomenal properties will share and the subjective character is identified with such a functional role. As long as we are concerned with the V-identity principle, all which is required is that, if phenomenal consciousness is sharp, so is the corresponding functional role identified with the subjective character.

Antony anticipates this and argues that a function is not sharp unless the realizer is also sharp. If N is the realizer of the function F, then F is sharp only if N is sharp.

> Suppose the system is in N, and that N realizes functional state F (=C). The system is thus also in F. Now assume N and F are correct. By gradually removing atoms from the brain we can generate a borderline case of N (see above). It will then be unclear whether that brain-state bears the same causal relations (actual and counterfactual) to inputs, outputs and other neurophysiological states that N did, so it will be unclear whether the system realizes F. We will thus have a borderline case of F as well. We are thus committed to this: If a property P realizes F, then a borderline case of P is a borderline case of F.

Antony seems to be explicitly endorsing the following principle:

(REALIZER)

> If P is a realizer of F, then borderline cases of P are borderline cases of F.

However, (Realizer) is false. Consider the case of the property being red. There are different shades of red and each of these shades is a realizer of the property being red. Consider for instance the property of being scarlet. If something is scarlet then it is red. But clearly borderline cases of being scarlet are not borderline cases of being red.

Antony does not have to endorse (Realizer), he requires a weaker principle to support his argument, something like:

(REALIZER-WEAK)

> If P is a realizer of F, then if P has borderline cases F has borderline cases.

16 Or a certain functional role and a certain constitution if one wants to avoid multiple realizability.

17 Once again if one wants to maintain that this change in functional role entails a change in the concept, I am happy to concede that. In that case the identification between N and phenomenal consciousness would be false, but what is relevant for materialism is that there is a physical property N* that can be identified with phenomenal consciousness, such that it is physical and is not vague.

(Realizer-Weak), however, is also false. There are cases of sharp functions with vague realizers. The property of *being an adder of two one digit numbers* seems to me to be completely sharp. This function can be realized by a child, by a calculator, etc. Both *being a child* and *being a calculator* are vague properties. There are indeterminate cases of *being a child* like a 12 years old boy. *Being a calculator* is vague as far as *being a neuron* is. I can run exactly the very same argument that Antony used to show it. I can start removing atoms from my calculator until I "reach borderline cases for concepts of many of those components [in this case the components of the calculator] (and their properties), and as a result borderline cases for [calculator] as well". If the function *being an adder of two one digit numbers* is sharp and the calculator is a realizer of this function then we have an example of a vague realizer of a sharp function, *pace* (Realizer-Weak). There will be a precise moment in the atoms removing process in which the calculator will definitely perform F and by removing one atom it will fail to perform F.

Another example against (Realizer-Weak) is that of a function F with only two possible realizers P_1 and P_2, both of them vague. That is, if determinately P_1 obtains or determinately P_2 obtains then determinately F obtains. If determinately neither P_1 nor P_2 obtains then determinately F doesn't obtain. Let's consider that every borderline case of P_1 is a borderline case of P_2. Consequently, if it is indeterminate whether P_1 obtains then it is indeterminate whether P_2 obtains and the other way around. However, P_1 or P_2 either determinately obtains or determinately does not obtain and therefore F either determinately obtains or it doesn't. F is a sharp function despite their realizers being vague.

If I am right, functional roles can be sharp without their realizers being sharp. That would allow for the identification of phenomenal characters with sharp functional roles avoiding panpsychism.

Antony could claim that regardless of the soundness of the (Realizer-Weak) principle, given the complexity required in a system for phenomenal consciousness, the function to be identified with phenomenal consciousness has to be vague. But this is either wrong or question begging. All we have to do is a functional analysis of this complex function, if our last day science can provide a sharp functional description of all the elements involved in it, then the complex function will be sharp. If we can provide a sharp functional description of neuron*, as the constitutive element of N* which is the realizer of the function F, then F will be sharp, even if the realizers of neuron* can be vague. Antony has no further argument, and insisting that phenomenal consciousness cannot be identified with F because it has vague realizers would just beg the question against the functionalist approach.

3.4 SUMMARY

In this chapter I have addressed objections to materialist theories of phenomenal consciousness that are based on Vagueness (VAM). This kind of arguments hold what I have called the V-principle. The V-principle maintains that the properties in an identity must share their borderline profiles. In general they maintain either that phenomenal characters are vague and physical properties are not or that phenomenal characters are sharp and physical properties are vague.

I have made a distinction between two different ways phenomenal characters can be vague: horizontally and vertically.

Horizontal vagueness relates to the qualitative character of the experience. An interesting argument in favor of the claim that phenomenal characters are horizontally vague is based on the non-transitivity of the *look the same as* relation. I have argued that materialists have the resources to accept that the relation *looks the same as* is non-transitive while resisting the claim that phenomenal characters are vague: the way the world looks depends not only on the phenomenal character of our experiences but also on the cognitive access we have to them. Furthermore, I have suggested that my view is to be preferred, for it provides the most natural explanation of how two experiences can be TC-distinguishable but not FS-distinguishable.

Vertical vagueness relates to the subjective character. I have reviewed some arguments in favor of the claim that the phenomenal characters are vertically vague and have argued that they are unappealing. There is no appealing reason for preferring Papineau's indeterminacy analysis of the harder problem to Block's epistemicist one. Furthermore, as Bermudez has argued, Papineau's position prevents any mind-body identity. On the other hand, I have maintained that Brogaard's argument based on gradable adjectives is supported by examples that confuse the notion of consciousness in play: phenomenal consciousness.

Finally I have presented Antony's argument against the materialist position. He argues that phenomenal consciousness is sharp but any plausible candidate which materialists have available for identification is vague. I have maintained, *pace* Antony, that functional roles can be sharp despite the fact that their realizers are vague. This blocks Antony's argument.

Materialism has nothing to fear from VAM arguments.

Part III

A NATURALIST THEORY OF CONSCIOUSNESS: SELF-INVOLVING REPRESENTATIONALISM

4 THE QUALITATIVE CHARACTER OF EXPERIENCE

In the first part of this dissertation I have presented and offered a reply to some arguments against materialism. The purpose of this second part is to offer a positive theory of phenomenally conscious mental states.

There is good empirical evidence to think that phenomenally conscious states are brain states. However, we also consider that beings lacking brains like ours may undergo phenomenally conscious experiences. What I consider to be the problem remaining, the problem that a theory of phenomenal consciousness has to address, is the problem of phenomenal properties. What are the properties, which some of my brain state have, such that when I am in these states I undergo a phenomenally conscious experience?

When I look at a red apple I undergo a phenomenally conscious experience. There is a *redness way it is like for me* to undergo the experience. This *redness way it is like for me* to undergo the experience is the phenomenal character of the experience. It is in virtue of its phenomenal character that the experience is the kind of experience it is and a phenomenally conscious experience at all. The phenomenal character of the experience determines that the experience is the kind of experience it is and a phenomenally conscious experience at all.

Experiences are a kind of mental state. I will be talking about the phenomenal character of the experience and about the phenomenal properties of the mental state. These uses are equivalent: the property of *undergoing an experience with phenomenal character* PC is identical to the property of *being in a state with phenomenal properties PP.*

This chapter and the next one try to explain what these phenomenal properties are; what are those properties that phenomenally conscious states have such that there is something it is like for its possessor to be in these states.

In the 1st chapter I have maintained that the phenomenal character can be decomposed into two components: the qualitative character and the subjective character (the *redness* component and the *for-me* component). The qualitative character distinguishes between different kinds of phenomenally conscious experiences. The subjective character makes an experience a phenomenally conscious experience at all. I will assume that:

> An experience has phenomenal character if and only if it has qualitative character and subjective character.

This claim is silent about the relation between qualitative character and subjective character. It may be that subjective character is a constitutive part of the qualitative character. In this case, having a qualitative character suffices for having a phenomenal character. Similarly, if the qualitative character is a constitutive part of the subjective character, then the subjective character suffices for having a phenomenal character.

Experiences are a kind of mental states: phenomenally conscious mental states. When S looks at a red apple and when she looks at a golf-course she undergoes two different phenomenally conscious

experiences; i.e. the subject is in two different mental states, M_1 and M_2. These mental states have different qualitative properties. Qualitative properties determine the differences in phenomenal character between phenomenally conscious mental states. It is in virtue of its qualitative properties that the experience I have while looking at a red apple is *as of red* and not *as of green* or *as of a symphonic concert*. Furthermore, both M_1 and M_2 are phenomenally conscious mental states. There is a property that these mental states have and that non-phenomenally conscious experiences lack. All phenomenally conscious experiences have a subjective character.

In 1.3.3, we saw that different philosophers maintain different positions with regard to the relation between phenomenal character and qualitative character. In this chapter, I want to remain as neutral as possible on the relation between phenomenal character and qualitative properties. I am not interested, in this chapter, in a theory of phenomenal consciousness that tries to explain what determines that certain mental states are phenomenally conscious; this will be the topic of the next chapter. In this one I will deal exclusively with the properties responsible for the differences in the phenomenal character of experience (*greenness, redness, etc*).

It is widely accepted that mental states depend on the subject's nervous system. What is controversial is whether phenomenal properties are intrinsic or extrinsic properties of mental states. More broadly, it is controversial whether phenomenal properties are intrinsic properties of the subject undergoing the experience or not. A classic example of an intrinsic property is the mass of an object. The mass of an apple depends exclusively on the apple. On the other hand, there are other properties that depend on the environment, like the weight of the apple, which depends, among other things, on the place where it is located. Weight and location are extrinsic properties. The distinction between intrinsic and extrinsic properties is nicely presented by Lewis (1983):

> The intrinsic properties of something depend only on that thing; whereas the extrinsic properties of something may depend, wholly or partly, on something else. If something has an intrinsic property, then so does any perfect duplicate of that thing; whereas duplicates situated in different surroundings will differ in their extrinsic properties. (ibid. pp. 111-112)

Though there are very interesting philosophical issues surrounding this distinction (see Weatherson (2006) for an excellent review) the intuitive idea is enough for my purposes. I will assume that we can abstract from the problems of a detailed characterization of this distinction for the discussion on phenomenal properties.

The internalist intuition maintains that phenomenal properties are intrinsic properties of the subject of the experience. According to this intuition, which is supported by our current knowledge of the brain, a microphysical duplicate of me undergoes the very same experiences as I do. I want to hold this internalist view about phenomenal properties. However, it should be said that this intuition has not gone unchallenged: extrinsic theories maintain that phenomenal properties depend on a certain relation between, say, an object and a subject. I will start this chapter by reviewing some of these extrinsic theories.

Some direct realist theories, for example, maintain that phenomenally conscious experiences have the phenomenal character they have in

virtue of an special relation between the subject and properties of objects. Two experiences of a subject differ in phenomenal character because the subject is related to different properties. Direct realism is motivated by the intuition that perceptual experiences are some kind of "openness to the world", what we do in perception is to, somehow, access the world. Direct realism faces serious problems given the existence of hallucinations, experiences in which there is no object of perception. In the first section I will provide a brief presentation of direct realism and my reasons for not endorsing this position.

Representationalism solves the problems of direct realism with hallucinations by appealing to the relation of representation. According to representationalism, the concrete phenomenal character of the experience is determined by the content of the experience or by the content of the experience and the functional organization of the mind. One of the most attractive reasons to hold representationalism is the transparency of experience: when we try to introspect the phenomenal character of the experience, we look "through" phenomenal properties and all that we do is to focus on the properties of the perceived object. Based on this observation, some philosophers have suggested that qualitative properties are representational properties: the intentional content determines the differences in character of two phenomenally conscious experiences. Representationalism is an appealing theory of qualitative character for materialists on the assumption that the relation of representation can be naturalized.

My purpose in this chapter is to present a theory according to which qualitative properties, the properties that determine the concrete phenomenal character of the experience, are representational properties of a particular kind.

In Section 4.2 I introduce representationalism and the transparency argument. I will also present some objections to the representationalist view. Representationalism has resources to deal with these objections, as I will argue. One of the objections I will present, the shifted spectrum objection, is especially pressing for those forms of representationalim that hold that representational properties are extrinsic properties of the subject. I will argue that narrow representationalism, the brand of representationalism I will embrace, can address the objection. According to narrow representationalism, the content of the experience[1] supervenes on the intrinsic properties of the subject: qualitative properties are intrinsic properties of the subject.

There are two questions that require further clarification:

1. What is the content of phenomenally conscious experiences such that it supervenes on the intrinsic properties of the subject?

2. In virtue of what does the relation of representation between what is represented (the content) and what does the representing (the vehicle of representation) hold?

I will address these questions in sections 4.3 and 4.4 respectively.

In section 4.3 I will provide a characterization of representational properties that respects the intuition, supported by empirical evidence, that phenomenal properties are intrinsic properties of the subject. This characterization should also address the problems of shifted spectrum

1 Unless otherwise indicated, 'content of experience' refers to the content of the experience that determines the phenomenology.

presented in the previous section. I will argue, following Shoemaker and Egan, that the correct characterization of the content of an experience with phenomenal character PC is the self-centered property that causes that experience in me *in normal circumstances*.[2] Any causal theory of content has to appeal to *normal circumstances* to account for the norm that distinguishes cases of veridical representation from cases of misrepresentation. An account of the representational content cannot be satisfactory until this apparent normativity is unpacked.

Representationalism appeals to the relation of representation to provide a theory of phenomenal consciousness. So, its plausibility as a materialist theory depends on the plausibility of a theory of mental content that is compatible with materialism. Theories of mental content try to explain the kind of relation that holds between what does the representing (the mental state) and what is represented (the intentional content) in such a way that it makes room for cases of misrepresentation. The most promising theories of mental content, teleological theories, appeal to the teleological notion of function to account for this norm: the content of a mental state is what the mental state has the function of indicating. Section 4.4 explores several of these theories of function. I will distinguish between etiological and non-etiological theories of function and argue that the former cannot be the satisfactory option.

4.1 DIRECT REALISM

Direct realist theories maintain that phenomenally conscious experiences of a subject have the phenomenal character they have in virtue of the special relation that holds between the subject and the object of perception. Two experiences of a subject differ in character because the subject is related to different objects with different properties. Direct realism is motivated by the intuition that perceptual experience is some kind of "openness to the world" (McDowell 1996). What we do in perception is to, somehow, access the world.

Direct realism is mainly interested in a theory of perception. According to direct realism, perceiving an object is an essentially relational state, of which the object perceived is a constituent; in other words, the perception is constitutively dependent on the object perceived.

I am interested in the phenomenal character of the experience, so the view that I will be considering here is as a direct realist position is, roughly speaking, the view that maintains that the phenomenal character of an experience of a red apple is constituted by the red apple. The redness or the roundness of your experience is nothing but a property of the red apple. The phenomenal character of your experience is given by a relation between you and the object of perception, the red apple. Phenomenal properties are these relational properties.

The phenomenal character of my visual experience of a red apple is constituted by a metaphysically primitive relation between the apple and my brain states. The object of perception (or properties of the object) itself is a constitutive part of the phenomenally conscious mental state.[3]

2 I will show in 4.3 how this rough characterization is compatible with the idea that two different subjects can undergo experiences with the same phenomenal character.

3 Much more needs to be said about this relation, but my reasons for not endorsing direct realist views are independent of these details, so I will abstract from them.

The main argument against direct realism is based on the possibility of illusions and hallucinations that are phenomenologically indistinguishable from veridical perceptions.

An illusion can be characterized following Smith (2002, p. 23) as "any perceptual situation in which a physical object is actually perceived, but in which that object perceptually appears other than it really is". Imagine that I am looking at a red apple but I see it as yellow. In this case I am suffering an illusion (for an example of an illusion see figure 1 in chapter 1). The problem for the direct realist is to explain the phenomenal character of this experience. The phenomenal character of the experience cannot be exhausted by the apple and its properties.

The case of hallucination is clearer. In this case, the subject is having an experience as of a red apple, but there is no mind independent object that the subject perceives. In the case of hallucination the subject is not related to any object. She is nevertheless undergoing a phenomenally conscious experience.

Direct realism maintains that the objects of genuine perception are mind-independent and the phenomenal character of a perceptual experience is constituted by these objects. Direct realism also accepts that illusions and hallucinations are possible.

Disjunctivism claims that these views are not inconsistent; they deny that genuine perception and subjectively indistinguishable hallucinations are mental states of the same kind. Disjunctivism denies what Martin (2004) calls the "common kind assumption" about perception:

> [W]hatever fundamental kind of mental event occurs when one is veridically perceiving some scene can occur whether or not one is perceiving.

What disjunctivists deny is the idea that what makes it true that these two experiences are phenomenally indistinguishable is the presence of the same fundamental kind of mental state in the case of perception and hallucination. Disjunctivism denies what Hinton (1973, p. 71) calls "the doctrine of the 'experience' as the common element in a given perception" and an indistinguishable hallucination.

Disjunctivism about phenomenal characters holds that some phenomenally conscious states are constitutively dependent on the object perceived and others not so.[4]

We can characterize disjunctivism as follows:

DISJUNCTIVISM: veridical experiences, illusions and hallucinations have different nature. Veridical experiences (and for some disjunctivist illusions) are relations between the subject and the object of perception, whereas hallucinations (illusions) have a different nature.[5]

I do not find disjunctivism to be an appealing theory of phenomenal properties. The problem for this position is that it seems to be ad-hoc. Let $\mathrm{PC}_{\mathtt{Veridical}}$ be the phenomenal character of a veridical

4 One can be a disjunctivist about perceptual states and not a disjunctivist about phenomenally conscious mental states by holding that a genuine perceptual state is a relational state between a phenomenally conscious state and the object of perception. However, that seems to concede the "common kind assumption".

5 Direct realists focus on the phenomenal character of perceptual experiences. They owe us a theory of the nature of phenomenal states like emotions or bodily sensations. Given that most direct realists are disjunctivists they could endorse any alternative account for these cases.

experience of a red apple RA. Let $PC_{hallucination}$ be the phenomenal character of an hallucination of RA. If $PC_{hallucination}$ can be indistinguishable from $PC_{Veridical}$, then disjunctivism is in trouble. The best explanation for the indistinguishability[6] of phenomenal character is that the phenomenal properties instantiated by the corresponding states are the same. If a property P, say being a certain brain state for instance, were what determines the phenomenal character of the hallucination ($PC_{hallucination}$), it is completely unclear why, if in the case of a veridical experience the subject also instantiates P, a further relation would be needed to account for the phenomenal character of the veridical experience.[7]

Alternatively direct realism could endorse something like the following:[8]

DENIALISM: denies that there can be hallucinations that are phenomenally indistinguishable from any veridical experience: hallucinations have a distinctive phenomenal character. Hallucinations and veridical experiences can have something in common but they are not phenomenalogically indistinguishable.[9]

It is an empirical question whether there can be cases of hallucination that are phenomenologically indistinguishable from veridical experience. I think that they are but, as far as I know, there is no conclusive empirical evidence to that effect. However, the current empirical evidence we have suggests that denialism is wrong. Some of the evidence and arguments that I will present against certain forms of representationalism will be evidence against any form of direct realism. I will make this explicit in a footnote when appropriate. Let me now focus on representationalism.

4.2 REPRESENTATIONALISM

Representationalism holds that phenomenally conscious mental states are representational states. A representational state is normally understood as one which is about, or represents, something in the world. There are other mental states with representational content, like those involved in thoughts, beliefs, desires, etc.

My belief that there is a red apple in front of me is about a red apple. It is not always true that when a representation represents something as being such-and-such, there is something which is actually such-and-such. There are cases of misrepresentation; for instance, Mateo believes that the Three Kings will bring him a lot of things on the 6th of January, despite the fact that, unfortunately, the Three Kings do not exist. Mateo's belief is false; it is a misrepresentation. Similarly, someone who thinks that phenomenally conscious experiences are a

6 In this case I am considering that the veridical and the hallucinatory experience are TC-indistinguishable. See 3.2.1.

7 Some disjunctivists (Martin (2004)) claim that the theory should remain silent about the phenomenal properties in the case of hallucination. I do not see any motivation beyond avoiding the commented problem for this quietist position.

8 This view has been suggested to me by Farid Masrou.

9 The direct realist that embraces this position owes us a theory of consciousness, because an hallucination is also a conscious experience. There should be something in virtue of which certain mental state is a conscious mental state at all. This is, however, not the topic of this chapter that concentrate in qualitative properties. Those properties are, according to the direct realist relations to objects or properties of the objects in the case of veridical experience and something different in the case of hallucinations.

form of representation can account for cases of illusion or hallucinations as cases of misrepresentation. My current visual experience is about a red apple, my auditory experience about the music from the CD, my olfactory experience about the coffee, etc. Hallucinatory experiences with the same phenomenal character as these experiences share with them, respectively, their intentional content.

The distinction we make between veridical and non-veridical experiences supports representationalism. An hallucination is non-veridical, whereas my current visual experience of the red apple is veridical. As Evans (1982) notes:

> We may regard a perceptual experience as an informational state of the subject: it has a certain *content* -the world is represented a certain way- and hence it permits a non-derivative classification as *true* or *false* (Evans, 1982, p. 226; emphasis in the original)

It seems that the experience one has when hallucinating a red apple has a red apple as its intentional object and that what the experience reports is false. That supports the idea that phenomenally conscious experiences are representational. Representationalism (as a thesis about the qualitative character) maintains that the satisfaction conditions of the experience, namely, there being a red object in front of the subject when she has an experience as of a red, exhaust its qualitative character. More precisely, representationalism holds that the content of the experience determines its concrete phenomenal character; i.e., there cannot be changes in the phenomenal character without changes in the intentional content. So, qualitative properties, the properties that determine that an experience is the kind of experience it is, are representational properties. We can therefore present representationalism as the following thesis.

> (Representationalism)
>
> Qualitative properties are representational properties.[10]

Although very different theories, as we will see, fall under the umbrella of this characterization, we can read representationalism as maintaining that in the case of a veridical experience of a red apple, the way it is like for me to see the red apple is constituted by the properties of the apple. The color of the apple, at least partially, constitutes the qualitative character of the experience; similarly to the position defended by the direct realist. But contrary to her (the *denialist* direct realist), the representationalist admits that there can be hallucinations that can be phenomenologically indistinguishable from a veridical experience. Representationalism maintains that in this case the intentional content[11] is exactly the same one than in the case of a veridical experience of a red apple and therefore both states have the same phenomenal character.

Representationalism, as presented above, is a thesis about the qualitative character; it maintains that differences in the character of phenomenally conscious experiences are determined by the content of the experience. It is silent on what makes a mental state a phenomenally conscious mental state at all.[12]

10 Notice that representationalism does not entail that every representational property is a qualitative property.

11 I will use intentional content and representational content interchangeably.

12 There is a stronger representationalist view, we can call it strong representationalism. Strong representationalism is not only a thesis about the qualitative character of the experience, but a thesis about the phenomenal character of the experience.

If qualitative properties are representational properties, then two experiences cannot differ in character unless they have different content. However, the experiences I have when I am looking at the red apple and when touching it are both about the form of the apple, but the visual and tactile experience clearly differ phenomenologically. Different philosophers have provided different replies to this worry; to understand their positions it will be useful to distinguish pure and impure representationalism.

Pure representationalism maintains that qualitative properties are pure representational properties. A pure representational property is the property of representing such-and-such. In the previous case, pure representationalism maintains that the properties that enter the content in the case of vision and in the case of touching are different.

Impure representationalism maintains that qualitative properties are impure representational properties. An impure representational property is the property of representing such-and-such in a certain manner. In the previous case, the same property is represented visually in one experience and tactilely in the other. Different manners in which a content can be represented can depend on the modality (visual, auditory, tactile, etc), on whether concepts are required (conceptual versus non-conceptual), on the corresponding propositional attitude (belief, desire, perception, etc). Impure representationalism requires not only a characterization of the content of the experience but additionally a characterization of the *manner of representation* (Chalmers (2004)). Such a characterization is commonly given in functional terms.

The Transparency of Experience

It is widely accepted that visual experiences are representational; they represent the world as being a certain way. What is controversial is whether these representational properties determine exclusively the qualitative character of the experience.

One of the most appealing reasons for endorsing representationalism is the so called *transparency of experience*. We normally "see right through" phenomenally conscious mental states to external objects and we do not even notice that we are in these states. When I look at my computer I am aware of a bunch of qualities. However I do not attribute these properties to my experience but to the computer (the *brightness* of the screen, the *blackness* of the cover, its having a rectangular form, etc.).[13]

This suggests that the properties that I am directly aware of when I have an experience are not intrinsic properties of the experience but properties of the representational content. The properties of the represented apple are constitutive properties of the phenomenal character of the experience.

(Strong Representationalism)

Phenomenal properties are representational properties.

The term 'representationalism' is often used to refer to this stronger thesis. As I have already stressed, in this chapter I am interested in the qualitative character of the experience; by representationalist theories I will always mean in this chapter theories that embrace (Representationalism).

13 In the next chapter I will precisify the transparency thesis and deny that the phenomenological observation supports the claim that the experience I have when I look at the apple merely represents the apple as having such-and-such properties. I will argue that the content of the experience is *de se*. When I undergo an experience I do not merely ascribe a property to the object of experience but I self-ascribe a certain property to myself.

Representationalism supplements this thesis with two further assumptions:

1. If phenomenally conscious states have relevant properties beyond the representational ones, they should be revealed by introspection.

2. Not even the most determined introspection ever reveals any such additional properties.

When we introspect the features of the experience all the features that we find are features of the representational content of the experience. When we introspect our experience, all that we find is its representational content, not any "mental paint" that bestows this content:

> Look at a tree and try to turn your attention to intrinsic features of your visual experience. I predict you will find that the only features there to turn your attention to will be features of the presented tree, including relational features of the tree 'from here'. Harman (1990, p.39)

One should be careful with the terminology here. Sometimes representationalists present their views by saying that phenomenal properties are identical to certain represented external properties, like the color red. This would be a categorical mistake, because phenomenal properties, as I have defined them, are properties of the mental state. As Chalmers (2004) notes this seems to be a mere terminological difference:

> This is a mere terminological difference, however. Dretske defines phenomenal properties ("qualia") as the properties we are directly aware of in perception, and concludes these are properties such as colors. This is quite compatible with the claim that phenomenal properties in my sense are representational properties, as long as one holds that one is directly aware of the represented property rather than the representational property. Once we make the relevant translation, I think that these representationalists' most important claims can be put in the terms used here without loss.

When I undergo an experience I am in a certain mental state. Representationalism maintains that the properties that this state has are not intrinsic properties of the mental state, but representational properties. Representationalism maintains that *having an experience with phenomenal character* $\mathrm{PC}_{\mathrm{RED}}$ is *being in a state that represents a certain property*, for instance physical redness; namely that phenomenal properties are representational properties.[14]

According to representationalism, the properties that my experience has are representational properties, phenomenally conscious states have the property of having a certain intentional content. When I introspect, I introspect the properties of my experience, phenomenal properties. The only properties that I am directly aware of in introspection are properties of the content of the experience, this suggest that phenomenal

14 For instance, what I call phenomenal properties is what Shoemaker (2001) calls qualia and what he calls phenomenal properties are properties of the content of the experience. I have chosen to call phenomenal properties to the former to stress that these are the properties I am interested in. I believe that phenomenally conscious states are brain states and my aim is to clarify what are the properties that these states have. Representationalism is the thesis that they are representational properties.

properties are representational properties. To say that my experience has phenomenal character PC_{RED} is to say that my experience has a certain property, say the property of being red, as is intentional content. What the representationalist denies is that there are intrinsic features of the experience beyond the representational ones. This more clear understanding is conceded by Harman himself:

> Can we become directly or introspectively (as opposed to inferentially) aware of those aspects of perceptual experience -the mental paint, etc.- that serve to represent what we perceive? I say we cannot.
>
> [L]et me insist that by "a representational feature of experience" I mean a feature of experience. My point was that we can be aware of such features of experience without being aware of the "mental paint" by virtue of which the experience represents what it represents. Harman (1996, p.75)

When I look at my apple, I undergo a phenomenally conscious experience, there is a redness way it is like for me to look at the apple. According to representationalism, having such an experience is, roughly speaking, being in a state that represents a property that the apple has (if the experience is veridical).

The same observation can be extended to the case of non-veridical experiences, like hallucinations and illusion. The phenomenal character of my experience is the same one when I see a red apple than when I hallucinate a red apple. There could be no object causing the experience; the *redness* in the hallucination, however, would be exactly the same property as the one involved in the veridical experience. In this case the red apple is a mere intentional object.

Representationalism generalizes these observations to other perceptual modalities and bodily sensations. As a result of these observations representationalism suggests that qualitative properties are identical to representational properties of a certain kind.

Representationalism is an appealing theory for materialists, on the assumption that the relation of representation can be naturalized, but not an uncontroversial one. In the next subsection I will discuss some arguments from the literature against representationalism. I will argue that there is a form of representationalism immune to these arguments and compatible with the view that phenomenal properties are intrinsic properties of the subject.

My purpose in this chapter is to present a representational theory of qualitative character that is compatible with the internalist intuition; i.e. two microphysically identical individuals will undergo the very same kind of experiences. For that purpose I will endorse the form of representationalism proposed by Shoemaker and refined by Egan and develop it into a naturalistic theory of qualitative properties.

Let me first face some objections presented against representationalist theories in general.

4.2.1 *Problems for Representationalism*

Arguments against representationalism tend to present cases where there is an apparent failure in the relevant relation between phenomenal

properties and representational properties. This section is divided in three subsections with different arguments against representationalism

In the first one, I present arguments to the effect that there are phenomenally conscious experiences that lack intentional content. I will argue that we have good reasons for thinking that the qualitative character of these experiences is also determined by the intentional content, and therefore deny that there are phenomenally conscious experiences that lack intentional content.

The second one presents the inverted spectrum objection. This objection holds that the intentional content can vary without a difference in the phenomenal character of the experience. I will argue that for that objection to succeed an inverted spectrum has to be metaphysically possible and we have good reasons for doubting that it is so.

The last one addresses objections that intend to show that there can be differences in the phenomenal character of the experience without differences in the content of the experience. I will first present an argument by Byrne that intends to show that these cases can be ruled out a priori. I do not find it compelling and I will argue why. Then, I present objections which are divided into two groups. The first one presents examples of experiences that allegedly have the same content but differ in character and I argue that representationalism has the tools to reply to these cases. The second one presents empirical evidence in favor of the claim that different subjects have color experiences that are systematically slightly different; that is, their spectrum is shifted. I will argue that this is a problem for some forms of representationalism but not for narrow representationalism, the kind of representationalism that I embrace. Introducing the details of this form of representationalism will be the task of the next section (4.3).

Phenomenal Character without Intentional Content

Some philosophers (Block (2003); Burge (1997); Loar (1990); Peacocke (1984)) have maintained that there is more to the phenomenology of bodily sensations (pains, itches, orgasms, etc.) and especially emotions (happiness, sadness, fear, etc.) and moods (depression, elation, anxiety, etc.) than what is represented. Block (2003), for instance, claims that there are introspectible features of the experience that are not representational, Block calls them 'mental oil'. Moods, orgasms and pains can illustrate this idea of *mental oil*. According to Block, they have a minimal representational content but are vividly introspectible.

Vision is plainly representational, for that reason it is not surprising that representationalism focuses on that modality. Perceptual experiences, in general, are good candidates for being representational. But, if representationalism is to succeed, restricting representationalism to perceptual experiences is not an option. First of all, this restriction seems to be ad-hoc and second, the qualitative character of experiences lacking content would remain completely unexplained: what explains the difference in character between the experience I have when I have headache and the one I have when I look at my apple?

I disagree with Block, I think that these kind of experiences do not support the claim that there must be something like 'mental oil'. Representationalism rejects this claim and suggests that every phenomenally conscious experience has intentional content. Bodily sensations do have intentional content. This can be easily shown in cases like pain. Pains are felt as being in a certain part of one's body and as if certain parts

are disordered in a certain way. Tye (1997) suggests that the qualitative character of pain is given by the content: "bodily damage or disorder" (ibid., p.113). This example can be extended to other bodily sensations.

Block seems, however, to be right in his complain that this content cannot exhaust the qualitative character of the experience; there seems to be something beyond this content,[15] something that we can introspect and that is also constitutive of the phenomenal character of experiences of pain: its affective phenomenology; i.e. it's awfulness, its urgency. This is the element that disappears in cases of pain asymbolia,[16] like in pains treated with morphine. If this component is not representational, representationalism is in trouble.

Martinez (2011) suggests that the affective aspect that some experiences have can also be analyzed in representational terms; but contrary to other cases such as perceptual experiences, the content is not *indicative* (that there is such-and-such) but *imperative*; something like: "don't have this bodily disturbance."

Emotions are often considered to be representational state. For instance Antonio Damasio (1995, 2000) have presents a collection of evidence in favor of emotions representing internal states of the body. This has been extended by philosophers like Prinz (2004) into compelling theories of emotions.

Supplementing these representational theories with imperative content, in order to account for the affective aspect of our emotions, or bodily sensations, seems to be a promising line of research, one that is left for future research. Suffice it to say, at this point, that emotions, moods, and body sensations are far from being a knockdown objection to representationalism.

An independent problem for representationalism related to the lack of content arises from naturalizing the notion of representation. I will come back to this issue in 4.4.1. Let me just sketch the idea. The most promising theories for naturalizing the content of mental states are teleological theories. People in philosophy of mind usually cast out the teleological insight in this way:[17]

> A mental state M represents such-and-such if and only if M has the function of indicating such-and-such.

If there could be a being that has phenomenally conscious experiences, but such that her mental states lacked any function, then representationalism would be false, for her mental states wouldn't represent anything.

Same Phenomenal Character, Different Intentional Content.

The main argument in favor of two experiences sharing their phenomenal character but differing in intentional content is the one based on the possibility of inverted spectrum.

15 Block (2003) further argues against the representationalist position by appealing to the qualitative character of orgasms.

16 Pain asymbolia is a condition that usually results from injury to the brain, lobotomy, cingulotomy or morphine analgesia. Typically, patients report that they have pain but are not bothered by it, they recognize the sensation of pain but are mostly or completely immune to suffering from it. See Grahek (2001)

17 If one consults the primary literature in teleosemantics, one will find plenty of complication, but this simplistic characterization captures the main insight and suffices for our purposes.

Block (1990, 2003) has presented several mental experiments to support the view that the intentional content can change without the corresponding change in the qualitative character. In the *Inverted Earth's* thought experiment Block invites us to imagine Inverted Earth, a possible planet where colors are inverted from ours, the sky and the sea are yellow, bananas are blue, ripe tomatoes are green, the grass is red, etc. Inverted Earth is as similar as possible to Earth in any other respect.

Block's thought experiment in (Block, 2003) is the following: you go to Inverted Earth wearing contact lenses to correct the differences in color in such a way that you do not notice any difference. What it was like for you to see a red apple before using the contact lenses in the Earth and what it is like for you to see an apple in Inverted Earth after wearing the contact lenses will be the same. When you arrive there, your experience misrepresents colors. You decide to integrate in the Inverted Earth society. Block's intuition is that, after a certain period of time, the experience you have while looking at the Inverted Earth's red grass is going to be about the red grass, so you are not misrepresenting anymore. However, the kind of sensation you have, the qualitative character of your experience, will remain the same. If this were so, representationalism would be false, for there would be a change in the intentional content without a change in the phenomenal character. Your experience is about a red object but the qualitative character is still *as of green*.

Most philosophers agree that, after sufficient time, the content of the experience does change. One can, nevertheless, resist that claim. Representationalists that maintain this line of reply could appeal, for instance, to a teleological theory of intentional content (Millikan (1989); Neander (1991); Papineau (1993)). No matter how much time you spend in Inverted Earth your experience will always be wrong. According to these teleological theories, the intentional content of your experience won't change because the content of your experience depends on your evolutionary history. However, I do not find teleological theories of mental content appealing for explaining the representational relation in the case of phenomenal properties. I will offer some objections to these theories in 4.4.

A more interesting reply is to deny that inverted spectrum is metaphysically possible. Representationalism is not an a priori thesis. Qualitative properties are not a priori entailed by physical properties as we saw in chapter 2. The lack of conceptual connection between qualitative properties and representational content of a certain kind explains the conceivability of inverted spectrum scenarios. The conceptual coherence required for the conceivability of such scenarios is not enough for supporting its metaphysical possibility, and this kind of possibility is what is required for proving representationalism wrong. The conceivability of Inverted Earth is perfectly compatible with representationalism as far as it is metaphysically impossible.

Furthermore, some philosophers have even doubted that it is conceivable given that our color space is asymmetrical. There are more perceptually distinguishable color shades between red and blue than there are between green and yellow. This would make red-green inversion impossible (Hardin (1997)).[18]

18 As we saw in chapter 1, Kalderon and Hilbert (2000) go a step further and argue that every quality space must be asymmetrical and so inverted scenarios are not possible.

Nevertheless, some forms of representationalism can be proven false even if an undetectable inversion is not metaphysically possible. All that is required for proving representationalism wrong is a *shift* in the spectrum. There is empirical evidence that suggests that the color experiences of normal subjects are slightly different. I will present this objection at the end of the next subsection. Let me first present other counterexamples in the same direction that can, I think, be handled by representationalism.

Same Intentional Content, Different Phenomenal Character

If there are two experiences that differ in phenomenal character but have the same intentional content then representationalism is wrong, for there will be changes in character without a change in the content of the experience. In this case, qualitative properties would not be representational properties.

Byrne (2001) has argued that such cases can be ruled out a priori, but I am not convinced by his argument. According to him, the content of a phenomenally conscious experience is the way the world seems to the subject of the experience. The idea that supports the argument is that if the way the world seems to a subject S doesn't change, then it cannot be that the phenomenal character of the experience has changed.

Consider S undergoing two consecutive experiences E_1 and E_2 that differ in phenomenal character. Assume that S is a competent subject in the sense of not having any cognitive shortcoming, in particular her memory is working properly. We can idealize S in such a way that she can perfectly remember the previous experience and compare its phenomenal character with the phenomenal character of the current experience. S will notice the change in the phenomenal character solely on the basis of the current experience and the memory of the previous one.

In that case, the way the world seems to her when she undergoes E_2 must differ from the way the world seemed to her while undergoing E_1. For otherwise, if the world seems exactly the same to S during E_1 and E_2, she has no basis for noticing a change in the phenomenal character.

The argument can be generalized *mutatis mutandis,* as Byrne shows, to the conclusion that experiences cannot differ in phenomenal character without differing in representational content.

However, as I have argued in 3.2.1, two experiences can in fact differ in phenomenal character without thereby differing in what Byrne understands as representational content: the way the world looks like to the subject. The world can look to us exactly the same in two experiences and nevertheless, those experiences may have different phenomenal character: in some cases, I can tell two experiences that look the same apart by comparison to a third experience, as we have seen in the previous chapter. Byrne considers this possibility in a footnote:

> [A subject] may have limited powers of discrimination (like us), it is a mistake to hold that she will always *know* that there is a change in phenomenal character: if the change is sufficiently small, she won't ... [T]he subject might not know (and hence not notice) that there is a change in phenomenal character when this is not one that the subject reliably discriminates –for short, when the change is negligible ... if two experiences differ non-neglibly in phenomenal charac-

> ter, they differ in content. Because negligible differences are intuitively borderline cases of differences in phenomenal character, we may strike 'non-negligible' and view the resulting supervenience thesis as true on a precisified but still perfectly reasonable sense of 'phenomenal character'.(Byrne, 2001, fn. 19)

In the previous chapter, I have offered some reasons for rejecting the intuition that negligible differences are intuitively borderline cases of differences in phenomenal character. I have claimed that *the way the world looks* depends on the cognitive access we have to the phenomenal character and not exclusively on the phenomenal character of the experience. The view that the way the world looks does not depend exclusively on the phenomenal character of the experience is to be preferred, because it can easily explain, as we saw, the fact that we can TC-distinguish two experiences which are not FS-distinguishable.

The proposal I made in the previous chapter could be rejected if one holds that the cognitive access is constitutive of the phenomenal character, but Byrne doesn't seem to endorse this view in his reply and in the next chapter I will offer further arguments against such a view. If I am right then negligible differences are not borderline cases of differences in phenomenal character and, *pace* Byrne, the phenomenal character can vary without a change in *the way the world looks*: in the case of two FS-indistinguishable experiences *the way the world looks* is the same but the phenomenal character of the experiences might be different.

The intentional content of an experience is what the experience is about. Representationalism holds that there is an interesting relation between phenomenally conscious experiences and intentional content (the relation of representation): qualitative properties, which, at least partially, determine the phenomenal character, are representational properties. It is not clear to me that the intentional content has to be *the way the world looks like* to a normal subject, at least if we hold on the folk use of *looks like*. Two experiences can be about different properties (two different shades of green for instance), have different phenomenal character, and nevertheless be such that the world looks the same to me when I undergo both experiences as we have seen in the previous chapter.

One could insist that there is a sense of the expression '*the way things look*' under which if two experiences are TC-distinguishable then things don't look the same after all. But in that case, Byrne's argument loses all its interest: two things look the same to me only if there is no way I can distinguish the experiences; namely, if they have the same phenomenal character.

I do not believe that Byrne's argument for the a prori impossibility of two experiences differing in phenomenal character without thereby differing in content succeeds. However, I do believe that qualitative properties are representational properties.[19]

Some philosophers have presented examples that intend to show that experiences can vary in character while having the very same content. I do not find them compelling and they have been rejoined by representationalists. As an illustration of this kind of objections, I will present some examples and the kind of reply representationalists offer.

19 Assuming that certain theories about the naturalization of the relation of representation are true, as we will see in 4.4.

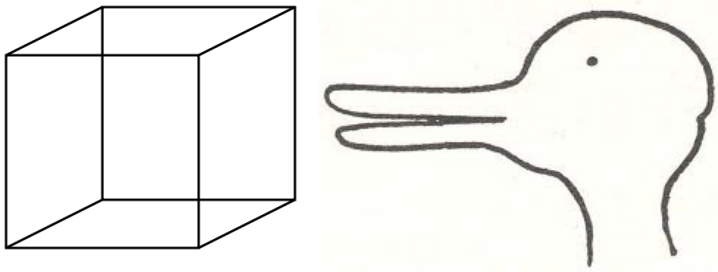

Figure 4: Ambiguous Figures: Necker Cube and Duck-Rabbit.

Peacocke (1984) presented three examples of experiences which seem to represent exactly the same property while differing in phenomenal character. In one of them, for example, he invites us to consider an experience that represents two trees of the same height and other dimensions but located at different distance from the observer. Peacocke suggests that "the nearest tree occupies more of your visual field than the more distant tree" (ibid, p. 12); i.e. there is a qualitative difference between the experience of each tree. Peacocke claims that both trees are represented as having the same height and this would be a problem for representationalism. However, representationalists have a very natural reply: one tree is represented as being further away from the observer than the other (Byrne 2001). In a more elaborated reply, Tye (1997) suggests that one of the trees subtends a larger visual angle from the subject's point of view, and this fact is itself represented by the visual experience.

Other interesting examples (Block (1996)) including cases of blurry vision, double images, etc. have been reasonably replied by representationalism (see Tye (2003a)). For instance, in the case of blurry vision, if one concedes that there is a phenomenological difference between seeing an object as being blurry, as when we look at a blurry painting, and blurrily seeing an object that it is itself non-blurry, then it is not enough to say that the visual experience represents the object as being blurry. Tye (2003a) argues that the differences can be accommodated in representational terms: in the first case vision represents the blurred edges as such, whereas in the second case, the problem is the insufficient information that vision tracks. Accordingly, Tye makes a distinction between non-veridically seeing a sharp object as blurry (misrepresenting the boundaries as fuzzy) and seeing the same object blurrily (not representing the boundaries in detail).

Ambiguous figures, such as a Necker Cube, the duck-rabbit picture, etc., have also been presented as problematic for representationalism.[20] In these examples an unchanging figure can give rise to visual experiences that differ in phenomenal character.

The representationalist's first reply, in line with the one given to the Peacocke's example, is to fine-grain the specific properties that are represented in each experience. For instance, an 'as of a duck' experience (see fig. 4) of the duck-rabbit will represent the property of being a bill without representing that of being an ear; the experience as of a rabbit will do the opposite.

20 For some suggesting examples see Nickel (2006) and Macpherson (2006) who offer a rich survey of ambiguous figures and rebut some possible replies that the representationalist could offer.

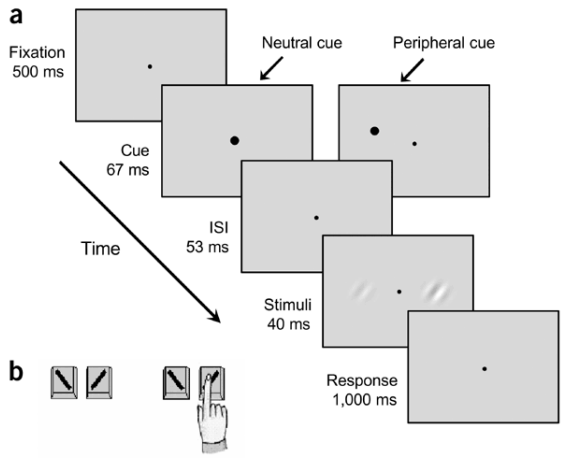

Figure 5: Carrascos' Paradigm for Measuring the Influence of Attention in Phenomenal Character.

Additionally, representationalists can appeal to different manners of representation in order to account for apparent cases of experiences with different character but same content. Let me present an objection to representationalism due to Block (2010) based on Carrasco's lab research on attention, to illustrate how this kind of reply would go.

It is well known that attention improves performance by increasing the accuracy and reducing the reaction time in tasks such as detection, discrimination, visual search, etc. Carrasco and colleagues have shown that attention alters the phenomenal character of the subject's experience. It alters saturation, contrast, spatial frequency, flickering rate, etc (Anton-Erxleben et al. (2007); Carrasco (2006); Fuller and Carrasco (2006)).

In a brilliant paradigm, Carrasco (2006) tested the subjective contrast perceived by the subject without asking the subject to rate their subjective experience, avoiding bias in the response while measuring the effect of attention in phenomenal character and performance.

In the experiment I am going to present, Carrasco used a common stimulus in psychophysics, an oriented grating whose luminance profile is a sinus. This kind of stimulus is called gabor patch. This gabor patches can be seen in figure 5 which illustrates the set up of the experiment. Subjects in the experiment are asked to fixate their gaze and attend to a point in the center. Two gabor patches will then appear. One of them has a fixed contrast and the other's contrast is modified randomly. The orientation varies randomly for both gabor patches.

In a first condition, subjects are asked to press a key with the orientation of the most salient gabor patch. If the more salient gabor is the one on the right, they have to use the keys on the right to indicate its orientation as shown in figure 5. The answer of the subject in this condition is compared to the answer of the subject in a second condition (see figure 5) where a cue appears and automatically captures attention. The cue can be neutral, and so it coincides with the fixation point, or peripheral. When the cue is peripheral, it automatically captures the attention to the side where it appears. This cue is uninformative: the relation between the position of the cue and the most salient gabor patch is random.

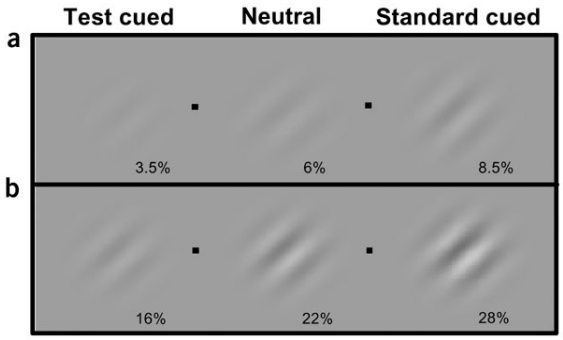

Figure 6: Attention Effect on Phenomenal Character

Figure 6 presents some of the results for high and low contrast. For example in the high contrast condition (b in figure 6) an attended gabor with a contrast of 22% looks like a gabor with a contrast of 28%. As a result of this experiment, Carrasco et al. have shown that subjects tend to perceive the cued gabor as more salient. Attention modifies the phenomenal character of the experience.

Block (2010) suggests that both experiences (attended and unattended) are veridical, that we have no reason for judging that either one is a case of illusion. But, if both of them are veridical then the properties we are representing must be different given that the experiences differ in phenomenal character. Nevertheless, the properties of the gabor patch remain the same, for the only thing that has changed is the subject's attention that has moved from the center to one of the sides. Block concludes that Carrasco's experiment shows that qualitative properties cannot be identified with the representational properties of the mental state.[21]

Representationalists can try to deny that both experiences are equally veridical. When the subjects attend to the gabor patch, the gabor is represented in more detail and so there is a difference in content. To back up this reply, the distinction psychologists make between prothetic and metathetic properties will be useful. Prothetic properties are properties with a meaningful zero value and inherent directionality such as saturation, contrast, spatial frequency, etc. For prothetic properties there is a gradable scale and it makes sense to talk about an increase in the information: no (zero) contrast, more or less contrast, more or less saturation, etc. On the other hand, there is no such a gradable scale for metathetic properties, like hue. Attention does not modify non-prothetic (metathetic) properties such as hue (Fuller and Carrasco (2006)). As Carrasco herself suggests, attention facilitates discrimination by modifying the perceptual capacity for prothetic properties. That seems to be a change in the content of the experience, an increase in the amount of information available to the subject.

For the sake of the argument, we can, nevertheless, accept that attention does not increase the amount of information and that the information processed in both cases is the same. If this is right, then attention facilitates discrimination by modifying the appearance. Carrasco's experiments do not pose a problem for impure representationalism. Impure representationalists can reply to Block by holding that, for

21 Block further argues against the direct realist position. The experiment shows a mental aspect that determined the phenomenal character beyond the properties of the observed object.

instance, the gabor patches are represented in a different manner when attended and not attended: attentively and not attentively. Representationalists should simply provide a functional characterization of this manner of representation.

There seem to be good reasons for believing that the strategy of the examples above does not show that phenomenal character can vary independently of content. If this is true, representationalism is safe.

Representationalism will be false, nonetheless, if the experiences of two normal individuals represent the very same property in the same manner but their experiences differ in character. There is empirical evidence to the effect that this possibility actually obtains.

SHIFTED SPECTRUM Block (2003, 2007c) has argued against representationalism providing empirical evidence that suggests that there is a variation in how colors appear to different normal subjects. If none of these subjects is misrepresenting then representationalism is jeopardized.

While looking at a red apple I have a phenomenally conscious experience with phenomenal character PC_{RED}. According to representationalism, the phenomenal character of my experience is determined by the property of representing red. The representationalist will be in trouble if the phenomenal characters of the visual experiences that two different subjects have while looking at the red apple could be different (granting that they are looking at the very same apple from the very same point of view under the same lighting conditions) and none of them could be said to misrepresent. They would be representing the same property in the same manner and nevertheless they would differ phenomenologically. There is empirical evidence that such cases are quite common.

Color perception in humans depends partially on particular light sensitive cells in the retina called cones. There are three kind of cones, each one responding mainly to light with a wave-length within a certain range. They are called accordingly *long*, *medium* and *short* cones. They have peak wave-lengths near to 564–580 nm, 534–545 nm, and 420–440 nm, respectively.

Lutze et al. (1990) have shown that there is a standard deviation in peak sensitivity of the cones of normal subjects of 1-2nm. This difference is very important, if we consider the variations in peak sensitivity between long and medium cones.

Furthermore, there a number of specific genetic divisions in the peak sensitivities in humans depending on sex, race and age (differences over 5nm!! Neitz and Neitz (1998)).

One should not be too quick to derive differences in the experience from differences in the peak sensitivity of the cones. Kraft and Werner (1994) study on the effect of aging in the visual system concludes that the visual system corrects certain alterations of the early stages of the visual system, in this case in the retina, but also that these corrections are insufficient.

> Age-related increases in ocular media density modify the spectral balance of broadband light reaching the retina. The visual system might compensate for this change, preserving the relative brightness of differently colored objects over the life span. Perfect compensation would require a function that is the exact inverse of the spectral absorption of

> the ocular media. Precise wavelength information is lost in the transduction process, however, so that an arbitrary spectral modification function cannot be constructed, and the visual system cannot exactly compensate for lenticular senescence ... sensitivity increase is spectrally broader and of lower magnitude than the ocular media density spectrum, increasing sensitivity inadequately where density is high and increasing it more appropriately where density is lower. Accordingly, brightness sensitivity remains constant (relative to the standard) at middle and long wavelengths but decreases at short wavelengths (420-480 nm) with increasing age. (ibid. p.1120)

There are further experiments that suggest that normal subjects can have a shifted spectrum. Block (2003) presents this evidence as follows:

> These differences in peak sensitivities don't show up in normal activities, but they do reveal themselves in subtle experimental situations. One such experimental paradigm uses the anomaloscope (devised in the 19th Century by Lord Rayleigh), in which subjects are asked to make two halves of a screen match in color, where one half is lit by a mixture of red and green light and the other half is lit by yellow or orange light. The subjects can control the intensities of the red and green lights. Neitz, et. al, 1993 note that "People who differ in middle wavelength sensitivity (M) or long wavelength sensitivity (L) cone pigments disagree in the proportion of the mixture primaries required" (p. 117). That is, whereas one subject may see the two sides as the same in color, another subject may see them as different–e.g. one redder than the other. When red and green lights are adjusted to match orange, women tend to see the men's matches as too green or too red (Neitz and Neitz, 1998). Further, variation in peak sensitivities of cones is just one kind of color vision variation. In addition, the shape of the sensitivity curves varies. These differences are due to differences in macular pigmentation, which vary with "both age and degree of skin pigmentation"(Neitz and Jacobs, 1986). Hence races that differ in skin pigmentation will differ in macular pigmentation. There is also considerable variation in amount of light absorption by pre-retinal structures. And this factor also varies with age.(ibid. p. 190)

From these evidences one can prove representationalism, as I have been presenting it, wrong. If there are shifted individuals; i.e., subjects that are shifted with regard to their visual spectrum, and none of them can be said to be wrong, then, *pace* representationalism, there will be a change in the phenomenal character without a change in the content of the experience.

According to representationalism, the qualitative character of the experience is exhausted by the intentional content of the experience.

(1) The content of the visual experience is constituted by the properties of the object perceived.

(2) The phenomenal characters of the visual experiences that two different normal individuals, S_1 and S_2, have while

looking at an object could be different (granting that they are looking to the very same object from the very same point of view under the same lighting conditions). Both experiences are veridical and about the same object and its properties.

(3) Both subjects are normal. There is no reason for establishing one subject perception as normal and claim that the other is misrepresenting.

∴ The difference in the phenomenal character of S_1 and S_2 is not exhausted by the properties of the external object.[22] Representationalism is wrong.[23]

If shifted spectra obtain then the phenomenal character can vary independently of the content and some representationalist theories are false.

One can complain that there is a degeneration of the early visual system due to age, and therefore the aged person misrepresents. The normal conditions under which the visual system is set up to function are not satisfied due to degeneration in the early visual system by aged people. In that sense aged people don't count as 'normal'. But such an alternative doesn't seem to be available in the case of gender or race.

Representationalists can accept that there is a difference in the qualitative character of the experience and explain the difference in terms of content –they can deny that the content of both experiences is the same. I will consider one possible externalist reply along these lines and show that it is not plausible[24] and then show how narrow representationalism can reply to it.

Representationalism can hold that the subjects, S_1 and S_2, represent a different set of properties or that the content of the experience of some subjects is more fine-grained than that of the others. Compare the experience of the two subjects while looking at a red apple from the same perspective with respect to color. Call the concrete shade of color of the apple RED_{34}. The representationalist can defend that RED_{34} is not a simple property; it can be decomposed into other properties ($RED_{34} = C_1 + C_2 + ... + C_n$). In this case, it could be the case that one of the subjects, or both, is not sensitive to one of the components of RED_{34}.[25] For instance, S_1 is not sensitive to C_3 and fails to represent this property. There would be a difference in phenomenal character explained by a difference in content. Nevertheless, in a certain relevant sense both S_1's and S_2's experiences are about RED_{34}.

This option doesn't seem completely plausible given the way the visual system functions and again it can empirically be proven false. A good candidate, according to our science, for being the component of RED_{34} would be light with different wave-lenghts. If this were the case, the proposal could be tested using light with a single frequency

22 This is also a refutation of the kind of direct realism considered in the first section.

23 Appealing to a functional notion as manner of representing is of no help here for both subjects represent the object in the same manner.

24 Something similar to this possibility is suggested by Tye (2002).

25 Someone could doubt that the experience can represent RED_{34} if it doesn't represent all of its components. The answer to this question will depend, among other things, on the details of the theory of mental content. I will concede that it is possible for the sake of the discussion.

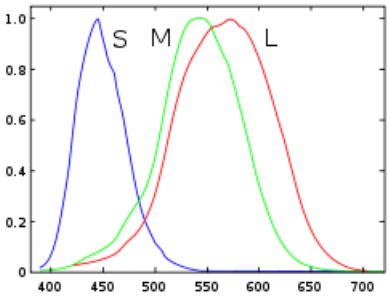

Figure 7: Normalized responsivity spectra of human cone cells, S, M, and L types.

component –so called delta signals.[26] Representationalists that hold on this reply would expect that the less sensitive subject cannot perceive certain signals with frequencies within the visual range; i.e., there would be certain properties that the less sensitive subjects cannot represent. I believe that this experiment would prove them false, as it is suggested by the current knowledge that we have.

We know that there are properties that normal subjects cannot come to represent. The audible frequency range goes from ca. 20Hz to 20KHz, but there are differences between individuals. A tone with a frequency of 19,9 Hz can be heard by some subjects and not at all by others. Something similar happens for visual perception. A typical human eye responds to wavelengths from about 390 to 750 nm. Thus, similarly, there would be colors that can be perceived by certain subjects but not by others. However, representationalists that want to hold on this rejoin will predict *gaps* in the middle of the visual bandwidth. If an illuminated circle with a single light component (in frequency) is presented to the subject and its frequency is modified from 390nm to 750nm in steps as small as necessary, the representationalist will predict that at certain frequencies the circle will disappear; the stimulus will not be visible at this frequency by the subject (in the example above, S_1 won't see anything when the only component of the stimulus is C_3). However, cones do not respond as an on/off switch, they have a Gaussian response profile.[27] The fact that there are cone cells responding to light frequencies within the visual bandwidth is a very good reason for believing that there are no such *gaps* in visual response.

Shift cases show that a certain form of representationalism is wrong: the kind of representationalism that holds that the representational properties that determine the qualitative character of the experience are extrinsic properties of the subject (wide representationalism). I am going to argue that there is a way of saving representationalism and the transparency observation (always assuming that there is a satisfactory theory of mental content compatible with materialism). In particular, I am going to argue in favor of narrow representationalism. According to narrow representationalism, the representational properties that account for the qualitative character of the experience are intrinsic to the subject;[28] i.e., if we fix the intrinsic properties of the subject,

26 This way the context dependency in color perception responsible for many illusions is also removed.

27 Figure 7 displays the normalized response profile of human cone cells (http://neurolex.org/wiki/Sao1103104164)

28 By narrow representationalism I refer to theories that maintain that the content of experience –the content of experience that determines the phenomenology– depends

the content is fixed. This kind of representationalism supports the internalist intuition: phenomenal properties are intrinsic properties of the subject.

In the next section I am going to motivate this brand of representationalism and provide a characterization of its content.

4.3 WHAT IS THE CONTENT OF EXPERIENCE?

Representationalism tries to account for the differences in phenomenal character of two experiences in terms of their representational content. According to representationalism, qualitative properties are representational properties. A representationalist theory of qualitative properties needs to provide a reply to two related questions:

1. What is the content of experience?
2. What is a representation? What determines the relation that holds between the content and the vehicle of representation?

In this section, I will address the first question. I will propose a theory which I think is the best shot for representationalism. Furthermore, representationalism depends on certain not totally clear assumptions related to the second question that I will spell out in the next section.

In the first place, I will distinguish narrow and wide representationalism and give some reasons for preferring narrow representationalism. Then I will review some alternative candidates to be the content of phenomenally conscious experiences.

Let me start presenting the arguments that support the idea that, in certain cases, representational properties are not intrinsic properties of the subject.

Putnam (1975) presented a convincing argument to show that meanings cannot be individuated just by internal states (semantic properties are not intrinsic properties of the subject) and Burge (1979) extended this argument to the conclusion that the content of certain propositional attitudes depends not only on the subject's internal states but also on her environment. Having a concept with semantic content C or having a propositional attitude toward content C* are not intrinsic properties of the subject.

To take a well-known example, consider *Twin Earth*, a planet which is identical to Earth, including microphysical duplicates of Earth inhabitants, except that in Twin Earth there is no H_2O but XYZ, a substance with different microstructure but with similar observable properties. In Twin Earth the colorless, odorless substance that fills up lakes is XYZ and not H_2O. The inhabitants of Twin Earth refer to their language as *English* and call XYZ *water*. Imagine Oscar, a competent English speaker, and imagine Twin Oscar, the dopplegänger of Oscar in Twin Earth, a competent *English* speaker. Oscar's and Twin Oscar's beliefs that 'water is wet' cannot have the same content, despite the fact that they share all their intrinsic properties, for Oscar belief is about H_2O whereas Twin Oscar belief is about XYZ. This example suggests that intrinsic properties do not suffice for individuation of semantic content or the content of our belief and desires. In order to have a mental state about

exclusively on the intrinsic properties of the subject. This claim is different from the claim that the content itself is intrinsic. Once my intrinsic properties are fixed, the content of my mental state is fixed. This is compatible with the fact that the truth maker, what determines that my mental state is correct or not, is something external.

water, the subject (or, according to some theories that we will see in 4.4.1, a sufficient number of the subject's ancestors) must have inhabited the right environment.

Along these lines, wide representationalism holds that representational properties are extrinsic properties of the subject. Representational properties depend on the subject's environment.

On the other hand, we have the strong intuition that phenomenal properties are intrinsic properties of the individual. What it is like to be me and my doppelgänger is the same. Somebody microphysically identical to me undergoes the same experiences as I do. This intuition is empirically supported. Nonetheless, this intuition seems to be in tension with the transparency of experience if representational properties are extrinsic properties. The following three statements seem to be incompatible:

1. Qualitative properties are representational properties.
2. Phenomenal properties (and therefore qualitative properties) are intrinsic.
3. Representational properties are extrinsic.

Particularly, in the case of color experiences a) and b) seem to be incompatible.

a) The phenomenal character of visual experiences is determined by the represented properties of the objects; colors, for instance.

b) The phenomenal character of a color experience is fully determined by the subject's intrinsic properties.

Anti-representationalists (Block (1990)) give up (1). They deny that phenomenal properties are identical to, or supervene on, representational properties. Anti-representationalists are not committed to deny that phenomenally conscious experiences have content. They deny that the phenomenal character of the experience is determined by its representational content. Anti-representationalism has the counter intuitive consequence that there can be phenomenal duplicates that do not share representational content; by having the same kind of phenomenally conscious experience they do not attribute any common feature to the object of the experience.

Wide representationalism (Dretske (1995); Tye (1997); Lycan (1996)) gives up (2). According to wide representationalism, phenomenal character does not depend exclusively on the subject's intrinsic properties; phenomenal properties are representational properties and representational properties depend on the subject's environment. In order to have certain content, the subject (or a sufficient number of ancestors as we will see in the case of some teleological theories of content) must have inhabited the appropriate environment, as we have seen in the case of Twin Earth cases. Wide representationalism gives up the internalist intuition and suggests that what it is like to be me may depend constitutively on factors that may be far away from me and in a distant past. That seems to me a hard pill to swallow, but bad digestion is not the only problem that wide representationalism has to face as we will see.

We have good reasons for believing that wide representationalism is false, as we have seen in the previous section. Two individuals can veridically represent the color of the apple and nevertheless differ

phenomenologically. If the content of the experience that determines the phenomenal character were colors, properties of the surface of the objects, then this would be a violation of representationalism, for this thesis holds that two individuals cannot vary phenomenologically without thereby varying in the representational property. Is there any alternative to save representationalism? I will argue that there is: narrow representationalism.

Putnam/Burge examples show that there are representational properties that are extrinsic properties of the subject but not that all representational properties are extrinsic. There is no argument that supports the claim that the intentional content that accounts for the phenomenal character of the experience requires the environment in order to be individuated.

The alternative to wide representationalism is narrow representationalism. According to narrow representationalism, the content of the experience supervenes on the subject's intrinsic properties. This doesn't mean that the content of experience is internal. It only means that if we fix the subject's internal states we thereby fix the content of the mental state, what the mental state is about.

In what follows, I will search for a correct characterization of the intentional content of the experience. The characterization I am looking for has to satisfy two desiderata: i) respect the internalist intuition and ii) provide a satisfactory explanation of veridical shifted experiences.

In the remaining of the chapter I will review two alternative candidates for being the content of the experience. The first one appeals, following Shoemaker's proposal, to appearance properties. The advocate of appearance properties holds that, although my visual experience might represent colors, the phenomenal character of the experience is determined by the property of representing appearance properties (Shoemaker (1994)). I will review some candidates to be these appearance properties and I will argue, following Egan, for a characterization of appearance properties according to which they are self-ascribed properties: functions from *centered worlds* to extensions. The content of the experience is *de se*. The second alternative is Fregean representationalism. According to Fregean representationalism, qualitative properties are identical to the property of representing a certain Fregean content, a mode of presentation. I will show that, under certain plausible considerations, there is no much difference between the brand of representationalism I argue for and Fregean representationalism. However, I will provide some reasons for preferring the Shoemaker/Egan proposal to Fregean representationalism.

4.3.1 *Appearance Properties*

Representationalism can be saved by supposing that whereas shifted subjects represent the object as being the same color, whereas there is a content they both share (wide content), there is also a representational difference with respect to other property that is attributed to the object by the experience (Shoemaker (1994)). Call these properties *appearance properties*. According to this view, qualitative properties are identical to the property of representing these appearance properties. In this subsection I consider some candidates to be these appearance properties.

Projectivism and primitivism.

Projectivism holds that appearance properties are properties of our visual field. Projectivism generally maintains that colors are nothing but these appearance properties. In this case there is no further property (wide content) beyond that of appearance.

Projectivist representationalism holds that color experiences attribute these properties to objects and that qualitative properties are identical to the property of representing these appearance properties. Boghossian and Velleman (1989) maintain that qualitative properties are properties of the visual field. Two intrinsic duplicates will share the visual field, and thereby their phenomenal and representational properties. Color experiences represent objects as having these properties.

A view in the vicinity is what Chalmers (2004) calls primitivism. Primitivism about colors suggests that colors are primitive intrinsic properties whose nature is revealed by color experiences. According to primitivism, colors are constitutively connected to phenomenal properties. An example of primitivism is Shoemaker's (Shoemaker, 1990) 'figurative projectivism'. Figurative projectivism maintains that qualitative properties are properties of the experience, but associated with each qualitative property there is a property that seems to us to be instantiated in the world; when the subject instantiates a qualitative property, she perceives something in the world as instantiating the associated property. Such a property is in fact not instantiated neither by the external object nor by the subject's experience. This view seems doubtfully naturalizable:

> In fact, the "secondary qualities" that enter into the intentional content of our experiences are never instantiated anywhere. They live only in intentional contents; in Descartes terminology, they have only "objective" reality, never "formal" reality. Shoemaker (1994, p. 24)

Different theories along these lines have been presented by Maund (1995); Holman (2002); Wright (2003).

Projectivism and primitivism views involve the attribution of properties to objects, properties closely connected to those of the experience itself. These views face the counterintuitive consequence that color experiences are massively illusory because objects do not have the properties our experiences ascribe to them. Whereas some philosophers accept this conclusion (Boghossian and Velleman (1989); Holman (2002); Maund (1995); Wright (2003)), a theory that does not have this undesired consequence is to be preferred. Such an alternative are dispositional properties.

Dispositional Properties

Dispositionalism maintains that appearance properties are dispositions to cause certain kind of experiences, experiences with certain phenomenal character (Shoemaker (1994)). I am perfectly aware that as I have just formulated it, the account seems circular. Let me please leave this issue aside for the moment for the sake of clarity. I will dispel any worry about the circularity of this proposal in 4.5 and in the next chapter.

As in the previous case, we can distinguish dispositionalism about colors, which maintains that colors are the previous appearance prop-

erties and those positions that suggest that appearance properties are different properties than colors.

Dispositional representationalism holds that color experiences attribute such dispositional properties to objects and qualitative color properties are the properties of representing such dispositions. Dispositional representationalism is a form of narrow representationalism if the phenomenal character of the experience depends exclusively on the subject's intrinsic states.

A satisfactory characterization of the dispositional properties involved here is complicated, as Egan (2006a) has pointed out. I will be following his analysis here.

Consider S_a and S_b, two shifted individuals. Let PC_{RED_1} and PC_{RED_2} be the phenomenal character of the respective experience they undergo while looking at a red apple under identical lighting conditions. Let A_{RED_1} and A_{RED_2} be the appearance properties their respective color experiences attributes to the apple when looking at it under the same lighting conditions. Egan holds that a characterization of appearance properties should satisfy certain principles, the ones below are a slight modification of some of these principles for the case of shifted spectrum:[29]

DIFFERENCE: S_a and S_b represent the apple as having a different appearance property; i.e. A_{RED_1} is a different property than A_{RED_2}.

CORRECTNESS: S_a and S_b both represent the apple correctly, when they represent it as being $A_{RED_1}$and A_{RED_2} respectively.

POS-SAMENESS: S_a can correctly attribute the same appearance property to an object O_1 than S_b correctly attributes to an object O_2, having O_1 and O_2 different color. [30]

INCOMPATIBILITY: Correctly representing something as having A_{RED_1} should be incompatible with correctly representing it as having A_{RED_2}.[31]

The first two principles seem to be non-negotiable. DIFFERENCE is required for saving representationalism. If qualitative properties are

29 Egan (2006a) includes two additional desiderata:

CONSTANCY: The appearance properties should be features that are had by things even when unobserved.

NOVELTY: The appearance properties are not colors.

For simplicity I will leave them asside. It will be trivial to see that the proposal satisfies these two desiderata.

30 This principle roughly corresponds to what Egan calls Sameness. Egan considers two inverted subjects Ernie and Vert and presents his SAMENESS desiderata as:

> SAMENESS: The appearance property that Ernie's visual experience attributes to Kermit is the same as the appearance property that Vert's visual experience attributes to, for example, cooked lobsters and ripe tomatoes. (ibid. p.501)

My POS-SAMENESS, is weaker than Egan's. POS-SAMENESS simply demands to the characterization of appearance properties to make room for the possibility of two shifted individuals correctly attributing the same appearance property to relevantly different objects.
It should be noted, as Egan does, that POS-SAMENESS and DIFFERENCE do not impose contradictory demands on appearance properties. DIFFERENCE requires that the appearance property that S_a attributes to the apple be different from the one S_b attributes to it. POS-SAMENESS requires that the appearance property that S_a's visual experience attributes to O_1 can be the same as the one that S_b's experience attributes to an object of a different color.

31 This principle corresponds to what Egan calls Contrariness.

identical to representational properties,[32] then the phenomenal character cannot change without a change in the representational content. If S_a and S_b differ phenomenologically, and A_{RED_1} and A_{RED_2} are respectively the representational content of their experiences, then A_{RED_1} and A_{RED_2} must be different properties. Otherwise there would be a difference in the phenomenal character without a difference in the content and therefore phenomenal properties would not be representational properties.

CORRECTNESS is also required for saving representationalism: appearance properties have been introduced precisely to reconcile representationalism with the possibility of shifted spectrum without misrepresentation. It has to be possible for S_a and S_b to represent the red apple correctly despite differing phenomenologically.

We think that two shifted subjects, S_a and S_b, can undergo experiences with the same phenomenal character, say PC_1, when looking at objects with different colors. For that reason we would like POSSAMENESS to be true. If S_a and S_b undergo experiences with the same phenomenal character, PC_{RED_1}, while looking respectively at two objects O_1 and O_2 that have slightly different color (and neither is misrepresenting), they should share a representational content. S_a represents O_1 as having A_{RED_1} and S_b represents O_2 also as having A_{RED_1}, that's why their experiences share the qualitative character.

Finally, INCOMPATIBILITY supports the idea that when S_a learns that the apple is A_{RED_1} he should learn that the apple is not A_{RED_2}: S_a learns that the apple doesn't look like the things that appear to be A_{RED_2}.

With these desiderata in hand we can analyze different candidates for appearance properties.

Shoemaker (2000) proposes as a candidate for appearance properties the following disposition:

> (Some)
>
> A_{RED_1} is the property of being disposed to cause experiences with phenomenal character PC_{RED_1} in some subjects.[33]

As Egan has noted, (Some) has different readings depending on whether the existential operator is read with modal force or not.

> (Some possible)
>
> A_{RED_1} is the property of being disposed to cause experiences with phenomenal character PC_{RED_1} in some *possible* subjects.
>
> (Some actual)
>
> A_{RED_1} is the property of being disposed to cause experiences with phenomenal character PC_{RED_1} in some *actual* subjects.

32 According to impure representationalism, qualitative properties are identical to impure representational properties. The manner of representation also plays a role in determining the qualitative character. I will continue just considering pure representationalism for the simplicity of the exposition. All the discussion can be extrapolated *mutatis mutandis* to impure representationalism.

33 Similarly, A_{RED_2} is the property of being disposed to cause experiences with phenomenal character PC_{RED_2} in some subjects. I will follow this nomenclature in the discussion.

(Some possible) doesn't seem to be acceptable. On the plausible assumption that necessarily coextensive properties are identical, it fails to satisfy DIFFERENCE. It is not unlikely that almost anything would be able to cause an experience with phenomenal character PC_{RED_1} in *some possible* subject. If this is true, almost everything would have A_{RED_1} because all that it takes to be A_{RED_1} is to be causally efficacious. The same happens with A_{RED_2} defined as the property of being disposed to cause experiences with phenomenal character PC_{RED_2} in some possible subjects. If this is true, then A_{RED_1} and A_{RED_2} are the same property on the previous assumption that necessarily coextensive properties are identical. Furthermore, (Some possible) makes no room for misrepresentation (a necessary condition for there to be intentional content), for all that it takes to be A_{RED_1} is to be causally efficient. Everything that causes PC_{RED_1} is A_{RED_1} and therefore PC_{RED_1} cannot misrepresent.

If we focus on (Some actual), we can see that it also has two possible readings depending on the force of the term 'actual'(Egan (2006a, pp. 505-506)):

> (Some actual_@)
>
> Something is A_{RED_1} at a world w if and only if it disposed to cause experiences with phenomenal character PC_{RED_1} in some subject that exists in the actual world (@).
>
> (Some actual_w)
>
> Something is A_{RED_1} at a world w if and only if it is disposed to cause experiences with phenomenal character PC_{RED_1} in some subject that exists in w.

(Some actual_@) cannot be the characterization we are looking for. It restricts the kind of observers to the actual world and therefore it cannot account for merely counterfactual shifted spectrum. If S_b is a merely counterfactual subject, he doesn't exist in the actual world, and so CORRECTNESS cannot be satisfied. S_b cannot correctly represent the red apple as A_{RED_2}, for S_b doesn't exist in @, an consequently the red apple is not A_{RED_2}. If the red apple is disposed to cause PC_{RED_1} in S_a and PC_{RED_2} in S_b, then the proponent of (some actual_@) is committed to say that S_b would be misrepresenting the red apple as being A_{RED_2}, for the red apple is not disposed to cause experiences with phenomenal character PC_{RED_2} in *any actual observer*.

On the other hand, the problem for (Some actual_w), according to Egan, lies on the characterization of the kind of properties we want to enter the content. Imagine two apples, $Apple_1$ and $Apple_2$, that have the same color. We want to hold that if they have the same color it should be possible that they appear the same. $Apple_1$ has A_{RED_1}, it is disposed to cause experiences with phenomenal character PC_{RED_1} in some subject that exists in the actual world. Imagine that $Apple_2$ exists only in worlds in which none of the subjects in the actual world exist or in worlds where there are no observers. In this case, according to (some_actual_w) it would not have A_{RED_1}.

In order to solve these problems we can focus on a certain kind of subjects. That would allow mere possible objects and kind of subjects to be taken into consideration.

> (Type T)

> A_{RED_1} is the property of being disposed to cause experiences with phenomenal character PC_{RED_1} in observers of type T.

One might complain that it seems weird to include the kind of observer as a part of the content of experience. Whereas it seems plausible that when I look at the red apple I represent it as having the disposition to cause PC_{RED_1}, it doesn't seem very plausible that I represent the apple as having the disposition to cause PC_{RED_1} in male observers with brown eyes for instance. This worry can easily be solved, I think, by using indexicals.

> (Indexical disposition)
>
> A_{RED_1} is the property of being disposed to cause experiences with phenomenal character PC_{RED_1} in *me (or in my type)*.

(Indexical disposition) does not require having any kind of knowledge about the conditions for individuating the relevant type.

(Indexical disposition) admits of two different readings: a *de re* reading and a *de se* reading.

> (Indexical disposition de re)
>
> A_{RED_1} is the property of being disposed to cause experiences with phenomenal character PC_{RED_1} *in me or my type (de re)*
>
> (Self-attributed)
>
> A_{RED_1} is the property of being disposed to cause experiences with phenomenal character PC_{RED_1} *in me (de se)*

In order to clarify the difference between (Indexical disposition de re) and (Self-attributed) let me properly introduce the notion of de se content (Lewis (1979)).

DE SE CONTENT I like the view about mental content according to which the role of mental states is to distinguish between different possibilities. The content of mental states are ways of dividing the space of possibilities. According to this view, what is relevant to the content is that it excludes certain possibilities. For instance, Stalnaker (1999) suggests the following:

> To say or believe [or perceive] something informative is to rule something out -to say or believe that some of the ways the world might have been are not ways that it is. The *content* of what one says or believes should be understood in terms of the possibilities that are excluded. (ibid. p.134; emphasis in the original)

According to this view, my belief[34] that Assange's arrest is a farce distinguishes between worlds that I take to be candidates to be actual; namely, worlds in which Assange's arrest is a farce, and worlds in

34 I start here talking about the content of propositional attitudes like beliefs, desires, etc. That may surprise some readers. However, I do it that way for the sake of the clarity of the exposition, because this is the origin of the discussion about *de se* contents. All that is relevant in this subsection is to clarify the notion of interestingly de se content and not whether it is the content of beliefs or experiences.

which it is not. The division in the logical space is made according to the corresponding proposition, in this case 'Assange's arrest is a farce'.

I will understand propositions as functions from worlds to truth values. The content of my belief is the set of worlds in which the believed proposition is true. In this sense, saying that the content of a mental state is a proposition is equivalent to saying that it is a set of possible worlds; namely those worlds where the proposition is true.

The set of worlds that constitutes the content is generally (if not always) determined by the attribution of properties to things. In the example of my belief the content is the set of worlds in which Assange's arrest has the property of being a farce.

For a given world, a property determines an extension. In the actual world (@), the property of being a farce determines an extension of all the things that are a farce in @. Assange's arrest is in the extension of the property being a farce in the actual world if, and only if, @ is a member of the proposition expressed by 'Assange's arrest is a farce.' In such a case, we can think of properties as functions from worlds to extensions.

In a similar way as we have defined propositions (possible-worlds propositions), we can define centered-world propositions as functions from centered worlds to truth values. The content of a mental state is de se if and only if it is a set of centered-worlds: the set of centered worlds in which the centered-world proposition is true.

If a possible world is a way the world might be, a centered world can be thought as a way the world might be *for an individual.* Centered worlds propositions do not just individuate a way the world could be, but also a certain logical position within this world. We can think of them as ordered pairs of worlds and individuals (<world, individual>). Egan (2006a) presents the notion of centered world as follows:

> A centered world is to a possible world what a map with a "you are here" arrow added is to an arrowless map. Centered worlds single out not just a way for the world to be, but a location within the world. They're best thought of as ordered pairs of a world and a center. There are different ways of picking out a center—the center could be, for example, a spacetime point, or an individual, within the world. Not much hangs on this decision, but it will be convenient for present purposes to take centers to be <individual, time> pairs. Some people talk about centered worlds, others about self-attribution of properties. Exposition is easier for centered worlds, but the same points can be made for self-attribution of properties. (ibid. p.518 fn. 34)

When I have a belief about myself, for instance, my belief is not well picked up as an attitude toward a proposition (understood as a set of possible worlds). As Egan (2006b) notes:

> Possible-worlds propositions do not cut finely enough - knowledge of, and belief about, possible worlds propositions can pin down which worlds I am in, but cannot pin down my location within that world.(ibid. p.106)

Possible-world propositions are determined by attributions of properties where properties are functions from possible worlds to extensions; centered-world propositions are determined by self-ascribed properties.

Self-ascribed properties are merely functions from centered-worlds to extensions. Egan calls this self-ascribed properties centered features to distinguish them from properties. He presents *being nearby* as an intuitively compelling example of the notion of centered feature. Which things are *nearby* does not depend exclusively on which world is actual, but also on where one is located "within this world"; i.e depends on a centered-world. If I am in Barcelona then the Parc Güell is nearby, if I am in Madrid, it is not.

The content of experience divides the space of possibilities. According to (Indexical disposition de re) the content is a property, a function from possible worlds to extensions. The way the world is suffices for the partition of the space of possibilities. On the other hand, according to (Self-attributed), the way the world is does not suffice for the division of the space of possibilities, additionally we require an individual.

When I have an experience with phenomenal character PC_{RED}, I attribute to the apple A_{RED}. According to the de re reading given by (Indexical disposition de re), this content is identical to the following: A_{RED} is the property of being disposed to cause experiences with phenomenal character PC_{RED} in Sebas (or in Sebas-type). This content divides the space of possibilities into worlds in which an apple is disposed to cause the experience in Sebas and worlds in which it is not. If the actual world is one of the former my experience is correct, otherwise it is incorrect. According to (Self-attributed)[35] a mental state is correct or not relative not only to possible worlds but also to individuals; the content of the experience are not properties, understood as functions from possible worlds to extensions. The content is *de se*. They are centered features or self-ascribed properties, functions from centered worlds, pairs world-individual to extensions.

(Indexical disposition de re) is an interesting proposal but it is not satisfactory as I will try to show. According to (Indexical disposition de re), when S_a looks at the red apple and has an experience with phenomenal character PC_{RED_1} she attributes to the red apple the disposition of causing PC_{RED_1} experiences in type-a subjects. Let me rename this property as A_{RED_a} to make it clear that it is indexed to type-a subjects. When S_b has a veridical experience with the very same phenomenal character PC_{RED_1}, S_b attributes to the object the property of having the disposition to cause PC_{RED_1} experiences in type-b subjects: A_{RED_b}. If a and b belong to different types, then A_{RED_a} and A_{RED_b} cannot be the same property, so POS-SAMENESS is not satified.

(Indexical disposition de re) also fails to satisfy INCOMPATIBILITY, representing something as A_{RED_a} is not incompatible with representing it as A_{RED_b}. Correctly representing the apple as A_{RED_a} does not rule out the possibility of representing it as A_{RED_b}. When S_a learns that the apple has A_{RED_a}, she does not thereby learn that the apple is not A_{RED_b}; namely, that it is not disposed to cause, say, PC_{RED_2} in S_b. It nevertheless satisfies something very close. When S_a learns that the red apple is disposed to cause experiences with phenomenal character

35 As Egan notes, some people talk about centered worlds, others about self-attribution of properties. Egan considers that exposition is easier for centered worlds and I am following his presentation, but the same points can be made for self-attribution of properties. This is the reason I called the proposal 'Self-attributed'. If we prefer to think in terms of self-attributed properties, (Self-attributed) maintains that in having an experience with phenomenal character PC_{RED} I attribute to myself the property of being confronted with an object that is disposed to cause a PC_{RED} experience in me.

PC_{RED_1} in observers of type-a, she thereby learns that it is not disposed to cause PC_{RED_2} experiences in observers of type-a. This desideratum in the vicinity of INCOMPATIBILITY is as far as we can go if POS-SAMENESS is not satisfied (for it guaranteess that A_{RED_a} and A_{RED_b} cannot be the same).

(Indexical disposition de re) faces an additional problem. Consider another kind of observer type-c. Assuming that necessarily coextensive properties are identical, then if type-c observers have experiences with phenomenal character PC_{RED_2} in the very same conditions as type-a observers have experiences with phenomenal character PC_{RED_1} then being disposed to cause experiences with phenomenal character PC_{RED_1} in type-a observers and being disposed to cause experiences with phenomenal character PC_{RED_2} in type-c observers will be necessarily coextensive and there won't be a difference in content between subjects of type-a and type-c who, *ex-hypothesi,* differ phenomenologically. This, as we have seen, leads to a denial of representationalism.

In order to satisfy POS-SAMENESS and avoid the problems of (indexical disposition de re) it is important that the same disposition can be attributed by shifted spectrum subjects when looking at different objects.

Properties, understood as functions from possible worlds to extensions, cannot do this job. We do not need a function from possible world to extensions, but a function from a centered-world (the dupla <world, individual>) to extensions. What we need is a function that, given a possible world and an individual in this world, delivers the extension. These are precisely the centered features or self-ascribed properties presented above. With this tool in hand, we can present a proposal that satisfies all the desiderata

> (Self-attributed)
>
> A_{RED_1} is the centered feature of being disposed to cause experiences with phenomenal character PC_{RED_1} in me (or in my type)

(Self-attributed) saves representationalism. Shifted individuals will differ in their representational content. A difference in the attribution of centering features is a representational difference. (Self-attributed) satisfies all the desiderata (Egan (2006a, p. 514)):

DIFFERENCE: If A_{RED_1} is the centered feature of being disposed to cause PC_{RED_1} in me, and A_{RED_2} is the centered feature of being disposed to cause PC_{RED_2} in me, then A_{RED_1} and A_{RED_2} are different centered features.

CORRECTNESS: S_1 and S_2 can correctly represent the red apple as having A_{RED_1} and A_{RED_2} respectively. The red apple is disposed to cause experiences with phenomenal character PC_{RED_1} in S_a and experiences with phenomenal character PC_{RED_2} in S_b.

POS-SAMENESS: S_a and S_b can correctly attribute the same centered feature to objects with different colors.[36]

36 It is important to note that the centered feature 'being disposed to cause experiences with phenomenal character PC_{RED_1} in me' is different from the property of 'being disposed to cause experiences with phenomenal character PC_{RED_1} in Sebas.' The latter corresponds to (indexical disposition de re) attribution, where the relevant disposition is indexed to individuals instead of types. The former corresponds to the attribution that (Self-attributed) makes.

In order to see that POS-SAMENESS is satisfied, consider two apples, $Apple_1$ and $Apple_2$ of different color. $Apple_1$ is disposed to cause experiences with phenomenal character PC_1 in S_a. A centered feature is a function from centered worlds to extensions, $Apple_1$ has the centered feature A_{RED_1}, if we introduce the actual world and S_a as arguments of the function we obtain an extension to which $Apple_1$ belongs to. This does not prevent that when we introduce the actual world and S_b as arguments in the very same function we will obtain an extension to which $Apple_2$ belongs to.

INCOMPATIBILITY: If S_a learns that O is A_{RED_1}, she learns that it is not A_{RED_2}. This is compatible with S_b learning that O is not A_{RED_1} by learning that it is A_{RED_2}.[37]

When S_a learns that O is disposed to cause an experience with phenomenal character PC_{RED_1} in herself, she thereby learns that O is not disposed to cause an experience with phenomenal character PC_{RED_2} in herself. This is compatible with S_b learning that O is disposed to cause an experience with phenomenal character PC_{RED_2} in himself, and thereby learning that it is not disposed to cause an experience with phenomenal character PC_{RED_1} in himself.

(Self-attributed) is appealing, but it has a fundamental problem. It makes no room for misrepresentation, whatever can cause the phenomenal experience in me has the appearance feature.[38] That seems to be wrong. When someone consumes LSD and hallucinates a red apple, we want to say that she is misrepresenting, we want to hold that the content of this experience is still as of a red apple and the subject is misrepresenting because the red apple didn't cause the experience. However, LSD has the disposition of causing an experience with phenomenal character PC_{RED} and according to (Self-attributed) LSD has the same appearance feature as the red apple. But clearly my visual experience doesn't represent LSD as having any appearance. This problem can be fixed by claiming that LSD is only disposed to cause experiences with the relevant sort of phenomenal character in a non-standard way. This way we can distinguish between the disposition LSD has, call it $PILL_{RED}$, and the disposition the apple has, A_{RED}. A_{RED} and $PILL_{RED}$ are different centered features. Objects that are A_{RED}, but not objects that are $PILL_{RED}$, are disposed to cause experiences with phenomenal character PC_{RED} in me in normal circumstances.

(Self-attributed*)

A_{RED_1} is the centered feature of being disposed to cause experiences with phenomenal character PC_{RED_1} in me *in normal circumstances*.

What is required next is a way of unpacking the *normal circumstances* that is compatible with materialism. That will be the job of section 4.4. Let me first review an alternative to these dispositional features as a candidate for the content of experience.

37 It might be useful, in order to illustrate how centered features satisfy these last two desiderata, to consider *being nearby*. When I learn that the parc Güell is *nearby* I thereby learn that it is not *far away*. This is compatible with David, who lives in Girona, learning that the park Güell is *far away* and thereby learning that it is not *nearby*.

38 Egan appeals to *non-deviant dispositions*, but does not elaborate on how this *non deviant dispositions* can be cashed out in terms compatible with materialism.

4.3.2 *Fregean Representationalism*

Chalmers (2004) suggests an alternative in the vicinity of the former proposal. According to him, qualitative properties are identical to the property of representing a certain Fregean content.

Frege distinguished between the sense and the reference of a linguistic expression. Taking an original example, the referent of the term 'Hesperus' is the planet Venus. The sense of 'Hesperus' is a mode of presentation, a certain condition the object has to satisfy for being the extension of the term. Something like: *being the object usually visible at a certain point in the evening sky*. According to Frege, the sense of an expression fixes the expression's referent. Venus is the object usually visible at a certain point in the evening sky and it is, therefore, the referent of the expression 'Hesperus'. Fregean contents are these modes of presentation.

Perceptual experiences involve, according to Fregean representationalism, modes of presentation of objects and properties. The mode of presentation will be the conditions the object or property has to satisfy to be the entity represented by my experience. For example, the phenomenal property that accounts for phenomenal character PC_{RED} would be the property of having the Fregean content that involves a mode of presentation such as *the property that causes experiences with phenomenal character* PC_{RED} *in me in normal conditions.* According to Chalmers, the representational content that accounts for the phenomenal character of the experience does not directly involve the property attributed by the experience:

> On this view, the relevant representational content does not directly involve the property attributed by the experience. It may well be that the experience attributes the property of redness to an object, and that redness is a surface reflectance property. The attributed property may enter the Russellian content of the experience, but it does not enter into the Fregean content. Rather, the Fregean content involves a mode of presentation. (ibid, p.174)

The content of an experience with phenomenal character PC_{RED} is RED where:

> (Fregean)
>
> RED is the property that has the mode of presentation: being the property that has the disposition to cause experiences with phenomenal character PC_{RED} in me in normal circumstances.[39]

Formally, modes of presentation are centered features: functions from centered possible worlds to extensions:

> [T]he Fregean content of a concept is a mapping from scenarios to extensions, and the Fregean content of a proposition is a mapping from scenarios to truth-values, where scenarios are maximal epistemic possibilities, or centered possible worlds. (ibid. p.172)

39 Chalmers presents it as something like 'being the property that causes experiences with phenomenal character PC_{RED} in me in normal circumstances.' but makes a dispositional reading of it. Chalmers acknowledges that his view "gives a key role to dispositional notions such as *normally causes phenomenally red experiences*."(Chalmers, 2004)

The very same centered features that figure as the content for (Self-attributed*) are presented in (Fregean) as modes of presentation. Both views give a central role to the centered feature of having the disposition to cause certain experience in me in normal circumstances. The difference is that whereas (Fregean) attributes to the red apple an intrinsic property, its color, with the dispositions serving as modes of presentation, (Self-attributed*) attributes to the red apple A_{RED}, an appearance feature.

The proponent of (Self-attributed*) is happy to concede that my experience also has a wide content beyond the appearance property. In that case, the color red is a good candidate for being the wide content. It seems natural to maintain that the experience represents a wide content in virtue of representing the appearance properties given by (Self-attributed*). It is not clear at all whether, in this case, there is any substantial difference between (Fregean) and (Self-attributed*).

Nevertheless, I feel inclined toward (Self-attributed*). The reason is that, though I agree that there is a function from centered features to the wide content, this function is not directly related to our experience as (Fregean) seems to demand.[40] Let me present an example to illustrate my claim.

Color experiences derive from the spectrum of light (distribution of light energy versus wave-length) interacting in the eye with the spectral sensitivities of the light receptors. I see no reason for believing that there is a unique property that can cause this in normal circumstances. It seems that, according to our physics, no unique physical property would be involved in color experiences. Even if we were color realists and accept that the apple has a certain color, probably an intrinsic property of its surface, whatever this property might be (surface reflectance?), it seems that this property would be different from the property involved in cases of color experiences due to light emission. Attributing a unique physical property in the case of color experiences seems to be wrong. Consider an hologram (light emission) that looks exactly like a red apple. This is really plausible if we consider the similarities between the color of objects in a 3-D film (light emission) and real objects (light reflection). We perceive the hologram *as red* and this perception is veridical, it is not a case of hallucination or malfunctioning.

Let's assume that the experience I have while looking at the apple and to the hologram are both correct and that I undergo experiences with the same phenomenal character. If there is a function from centered features to wide content and the wide content is a physical property of the object then the apple and the hologram should share a physical property, but it seems that they do not. The hologram shares with the apple an appearance feature (they both are disposed to cause the experience in normal conditions) and it is this feature the one that determines the qualitative character of the experience. Of course, there is a sense in which the experience is about the apple and its color, but

40 A detailed analysis of the relation between these two positions is a very interesting topic that I hope to work on in the future.
I prefer to remain neutral on this relation between narrow and wide content; I am interested in an account of the differences in character between phenomenally conscious experiences and only narrow representational properties are relevant for this purpose. Dispositionalism is silent on the relation between narrow and wide content.

this content seems to be irrelevant for the qualitative character of the experience.[41]

Consider another example. When I taste a banana I undergo a certain experience; my experience has a certain qualitative character, call it PC_{banana}. For (Fregean) my experience represents an intrinsic property of the banana, because it is the property that normally causes PC_{banana} in me. Imagine that a chemist creates a molecule that tastes exactly like a banana, but has nevertheless nothing in common –microphysically speaking– with the property responsible for the taste in a banana, namely BANANA. The property responsible for the taste of the flavor is a different property XANANA. However, when I taste the flavor I instantiate *PCbanana*. (Fregean) holds that my experience is not veridical because XANANA is not the property that normally causes *PCbanana* in me.[42] According to (Self-attributed*) both experiences are correct, BANANA and XANANA are both disposed to cause PCbanana in me in what intuitively are normal circumstances and, in fact, I attribute to them the same feature.

Preferences for (Self-attributed*) or (Fregean) would probably depend on the readers views about secondary qualities and her intuitions about cases as the ones presented above. Though I have made clear my preferences and the reasons for them, nothing of what I will say next depend on which of the two is the correct one.

Both, (Self-attributed*) and (Fregean), appeal to normal conditions to characterize the content of the experience. This notion is normative: it divides experiences into correct and incorrect ones. What is required next is a materialist compatible way of unpacking the *normal circumstances*. This is the target of the next section.

4.4 WHAT IS A REPRESENTATION?

As I have already mentioned, a theory of mental content has to clarify two questions:

1. What is the content of a representation?
2. What is the relation that holds between the content of the representation and the vehicle of representation?

Representationalism holds that the particular phenomenal character of an experience E, is determined by the content of the experience E. Qualitative properties, the properties in virtue of which my experience has a PC_{RED} character and no other character are representational properties. A subject S undergoes an experience with phenomenal character PC_{RED} only if S is in a state that represents A_{RED}; i.e. being in a mental state that represents A_{RED} is a necessary condition for having an experience with phenomenal character PC_{RED}. Whether or not all states that represent A_{RED} are phenomenally conscious states is left open for discussion in the next chapter.

The prospect of the representationalist approach as a materialist theory of phenomenal consciousness depends on having a reply, in terms compatible with materialism, to the following two questions:

41 An example where the 'emitted property' constitutes the wide content and not a 'reflectance property' is the case of a traffic light and a visually indistinguishable (from a certain point) photography of the traffic light.

42 This is compatible with phenomenal properties being intrinsic properties of the subject because qualitative properties are identical to fregean representational properties and the mode of presentation of BANANA and XANANA is the same one.

1. What is A_{RED}?
2. What is the nature of the relation between A_{RED} (the content of the representation) and M (the vehicle of representation)?

With regard to the first question, I have argued in the previous section in favor of an internalist view of phenomenal properties. Representationalism and internalism can be made compatible by appealing, following Shoemaker/Egan, to appearance properties. A_{RED} is the centered feature of being disposed to cause PC_{RED} experiences in me. This is not enough for individuating A_{RED}; the reason, as we have seen, is the following: when I see the red apple I have a phenomenally conscious experience with phenomenal character PC_{RED}. I thereby attribute to the apple A_{RED}, the disposition of causing PC_{RED} in me. When I take LSD or when a mad scientist manipulates my neurons I hallucinate a red apple. The apple, the LSD and the mad scientist have the disposition of causing experiences with phenomenal character PC_{RED} in me. We don't attribute A_{RED} to the LSD nor to the mad scientist. So, there are cases in which attributing A_{RED} to the cause of the experience is correct and cases in which it is not. We want to say that the LSD or the scientist caused the experience in the wrong way. Therefore, being disposed to cause PC_{RED} in me does not suffice for being A_{RED}, it has to be disposed to cause the experience *in normal circumstances*. This normal circumstances offer a criterion for distinguishing A_{RED} things from other things that are disposed to cause the experience. *Normal circumstances* is a normative notion that need to be unpacked in terms compatible with materialism, if representationalism as a naturalist theory is to succeed.

Let's focus now on the second question.

The most promising theories for naturalizing the content of mental states are teleological theories. The teleological insight is usually cast out, as claiming that a mental state M represents C if and only if M indicates C in *normal conditions**; where the notion of indication can be spelled out as a causally grounded tracking of information (Martinez (2010)):

> M indicates C if and only if:
>
> i) M tracks information about C: $P(C \mid M) > P(C)$
>
> ii) The difference in probabilities in i) is causally grounded

M indicates plenty of things; but we don't want to maintain that M represents all the things that it indicates, because all that it takes to be indicated by M is to correlate with it and there being a causal ground for this correlation. M represents exclusively those entities that it indicates in *normal conditions**. Someone might think that these normal conditions are different from the conditions that distinguish things that are A_{RED} from other things that are disposed to cause the experience (normal conditions); I will suggest that they are not.

We need to unpack the normative notion *normal conditions**. There are cases in which being in M is correct and cases in which it is not. It is correct when *normal conditions** obtain, and incorrect when they do not obtain. *Normal conditions** is a normative notion. According to the most promising theories in the project of naturalizing the content, teleological theories (Millikan (1984); Neander (1991); Papineau (1993)), this notion can be unpacked by appealing to functions. Dretske (1988) maintains

that a representing system is one that has the function of indicating that such-and-such is the case, being such-and-such the intentional content:

> [A representing system is] any system whose function it is to indicate how things stand with respect to some other object, condition or magnitude. (ibid. p. 52)

According to this idea, we can understand what it means that a mental state M represents a certain content C.

> (Representation)
>
> A mental state M represents C if and only if M has the function of indicating C.

The relation that holds between the representation (M) and the represented (C) has to make room for cases of error in which the contentful state is a misrepresentation. Cases of misrepresentation are cases in which the system is malfunctioning, cases in which M is activated when it should not. Hence, the required notion of function has to account for such normativity if the relation of representation is to be explained in terms of the function of indicating.

I suggest that by appealing to the function of the mental state M we can also unpack the normal conditions in A_{RED}. When I undergo a phenomenally conscious experience with phenomenal character PC_{RED} I am in a certain mental state M. This mental state represents A_{RED} things; that is to say that M has the function of indicating A_{RED} things. M is correct when its activation is caused by red apples or tomatoes and it is not correct when its activation is caused by LSD or by a mad scientist. The function of M determines when its activation is correct and when it is not and as a result its activation is correct when it is caused by the dispositional property of the apple and not when it is caused by the disposition of the LSD. We can appeal to the function of the mental state to discriminate two kinds of dispositions. The apple has, whereas neither the LSD nor the mad scientist have, the disposition to cause the experience in *normal conditions*. In this case, *normal conditions* and *normal conditions** are identical:

- M represents A_{RED} in virtue of M having the function of indicating A_{RED}.
- The apple has A_{RED} because it can cause the activation of M in the conditions fixed by the function of the state.

> M being active is correct if and only if it is caused by things that have the feature that in normal circumstances causes M, namely A_{RED}. M represents A_{RED} and these normal conditions are unpacked by appealing to the notion of function.

What is required next is a satisfactory account of the *function of a system* that is compatible with materialism; namely, a characterization of function that accounts for the intrinsic normativity in terms that are compatible with materialism. In the remaining of this section I will try to clarify this notion.

In the last twenty-five years there has been a renewed interest in philosophy of mind in functions and functional explanation with the hope that the notion of biological function would contribute to an

account of mental content. This account is fundamental, for instance, for addressing multiple philosophical problems for naturalistically oriented semantic theories. In our case, its interest lies on naturalistic theories of phenomenal consciousness that maintain that differences in phenomenal character are determined by the content of the mental state.

The interesting notion of function for mental content is one according to which the function of a trait is not necessarily something the trait does, but rather something that it is supposed to do. For example, it is said that the function of kidneys is to filter toxins and waste products from the blood, even in the case of someone suffering from renal insufficiency. The function attribution is normative:

> Function attributions are, in other words, not descriptive (they do not tell us what is the case) but normative (they tell us what should be the case). From this point of view, the main task of a theory of function is to explain how this norm arises in biological contexts. Wouters (2005, p. 124)

Theories of functions in biology can be grouped into two categories: etiological theories and non-etiological theories, depending on whether or not the function attribution relies on the causal history of the trait. For etiological theories, functions are selected effects, the effects for which the trait was selected for in the past. On the other hand, non-etiological theories hold that the function attribution is independent of the causal history of the trait. I will use the expression 'etiological function' to refer to the former and 'non-etiological function' to refer to the latter.

In the first subsection, 4.4.1, I present etiological theories of function. I will reject them as providing us with a satisfactory notion for explaining the representational content in the case of phenomenal consciousness. First I will show that if we appeal to an etiological theory of function, the resulting representational properties are extrinsic properties. Then I will present three original arguments against representational theories that rest on an etiological theory of function. In the second subsection, 4.4.2, I present non-etiological theories of function and maintain that there are good reasons for believing that a non-etiological notion of function will account for the required normativity. I will conclude the section with a dilemma, either non-etiological theories of function can satisfactorily account for the required normativity or we better give up representationalism.

4.4.1 *Etiological Functions*

The mainstream answer to the problem of normativity suggests that the function of a function bearer is the reason why the bearer is there. Etiological theories about functions maintain that the function of a trait should be identified with the reasons for the trait's existence. They follow Wright's ideas, previously voiced by Ayala (1970), who proposed the following definition Wright (1976, p. 81):

> The function of X is Z if and only if:
>
> i) Z is a consequence (result) of Xs being there,
>
> ii) X is there because it does (results in) Z.

This etiological notion can be unpacked in terms compatible with materialism. In the case of artifacts by appealing to the intentions of the designer: the function of my computer's cooler is to lower the CPU's temperature, because that's the reason why the designer placed the cooler in the CPU. In the case of biological traits, it can be unpacked by appealing to what the trait has been selected for, where selection is understood as natural selection or some other natural process of selection:[43] the function of the kidney is to filter blood because filtering blood is what the kidney has been selected for.

The function of a trait depends, according to etiological theories, on its causal history and past selection for traits of that type. That allows us to explain cases of malfunctioning for traits that have never performed their function. A cooler malfunctions when it doesn't decrease the temperature of the CPU. The cooler has this function even if it has never decreased the temperature of the CPU, because decreasing the temperature is what it was designed (and included in the computer) for. Similarly in the case of the kidney.

In the case of mental content, some teleological theories make use of this notion of etiological function. This notion is what plays the most important role in content individuation. Teleosemantic theories of mental content share the idea that the normativity of representation is given by etiological functions. Different theories of function based on the etiological notion of function have been presented by Millikan (1984, 1989); Neander (1991); Papineau (1993).

A representationalist who embraces an etiological theory of function (I will call this position etiological representationalism) will defend something like the following:

(Etiological Representationalism)

1. Qualitative property Q is the property of representing a content C.

2. Mental state M represents C because it has the etiological function of indicating C.

With a bit more of detail:

A mental state M of a subject S represents C if and only if:[44]

1. M has tracked information about C in a sufficient number of S's ancestors.
2. M tracking information about C has contributed positively to the fitness of S.
3. The conditional probabilities implicit in 1 are causally grounded.

43 A proper understanding of the notion of etiological function that is compatible with natural selection requires substituting (ii) with:

(ii') X is there because tokens of X's type did Z (in the past)

because selection does not depend on what the trait does, but on what it did.

44 The proposal presented here has the problem of not being able to individuate a unique entity that M has the function to indicate. This is known as the indeterminacy problem. For a detailed analysis of the problem and an etiological proposal as a solution see Martinez (2010, chapter 1). Nothing of what I say here rests on the details that solve the problem. The proposal I am discussing here intends merely to capture the insights of the etiological understanding of function.

There are, nevertheless, reasons for rejecting representationalism if it has to appeal to an etiological notion of function. I will first clarify that for etiological representationalism phenomenal properties are extrinsic properties of the subject, then I will present three original arguments against etiological representationalism.

Phenomenal Properties are Extrinsic Properties

The first problem, if we have to appeal to etiological functions to naturalize *normal conditions*, is that the content of the mental state will depend on the environment that the subject's ancestors have inhabited; i.e., the environment in which the state has been selected for.

According to etiological representationalism, the content of a mental state M is what it was selected for indicating.[45] Two microphysically identical individuals might differ phenomenologically if their ancestors had inhabited different environments and their respective states were selected for indicating different things. If qualitative properties are identical to these representational properties they are not intrinsic properties of the subject. So, phenomenal properties would be extrinsic properties of the subject.

This goes against the internalist intuition that microphysically identical subjects cannot differ phenomenologically. What it is like to be me would depend constitutively on factors that may be far away from me and in a distant past. That seems to be a hard pill to swallow, but maybe something we have to learn to live with.

Unfortunately for representationalists that appeal to etiological functions, even if they bite the bullet and accept this counterintuitive consequence, they face further problems.

Swampman

According to etiological representationalism, a trait has a function only if it (tokens of its type) has been selected for. When the trait appears for the first time it lacks a function. So, when the mental state M[46] appears for the first time in a subject (in the evolution) it doesn't represent anything. If phenomenal properties are representational properties and the relation of representation is explained by appealing to an etiological theory of function, then Swampman's state doesn't have the function of indicating anything: its state has not been selected for and, therefore, there was anything it was like for that subject to be in M; i.e., M lacks phenomenal properties. That's problematic, let me elaborate:

In his paper "Knowing One's Own Mind" Davidson (1987) presented a philosophical character, Swampman, to show the relevance of causal history for reference. This character is very useful to illustrate the problem. Davidson introduced Swampman as follows:

> Suppose lightning strikes a dead tree in a swamp; I am standing nearby. My body is reduced to its elements, while entirely by coincidence (and out of different molecules) the tree is turned into my physical replica. My replica, Swampman, moves exactly as I did; according to its nature it departs the swamp, encounters and seems to recognize my friends, and appears to return their greetings in English. It

45 Properly speaking the content of a mental state M is what traits of M's type were selected for indicating.

46 Obviously, by this I mean a token of the type M.

> moves into my house and seems to write articles on radical interpretation. No one can tell the difference. (ibid, p.19)

Intuitively, Swampman would have the very same phenomenally conscious experiences Davidson would have had. Yet, etiological representationalists are committed to deny that Swampman has phenomenally conscious experiences because lacking an evolutionary history, he lacks any function. If having a function is a necessary condition for the mental state to represent something, then Swampman's mental states do not represent anything. If phenomenal properties are representational properties, then Swampman lacks any phenomenal properties. Swampman has no phenomenally conscious experiences at all.

Etiological representationalists can complain that we lack clear intuitions in weird cases as that of Swampman (Millikan (1996)) or face up to it and claim that our intuitions are simply wrong (Dretske (1995)).

Some philosophers (Millikan (1996)) reject such a fanciful thought experiment. It is so far away from anything we can really take in, she argues, that our intuition about it can hardly show anything about our concepts. I think this is wrong and, nowadays, the exotica can become reality. In order to increase the size of the bitter pill that etiological representationalism has to swallow, let me present an original variation of Swampman's story.

Genetic engineering makes it possible to create individuals completely outside the evolutionary history. DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups. Attached to each sugar there is one of four types of molecules called bases – Adenine (A), Thymine (T), Guanine (G) and Cytosine (C).

Having a map of Davidson DNA, it is possible to create a DNA duplicate in the laboratory. This chain is introduced in a cell with the basic proteins to express this genome, a totipotent stem cell.[47] The conditions for its reproduction are guaranteed and some months later Swampman-Dolly is born. Swampman-Dolly lacks evolutionary history and therefore any function. No one seriously thinks that Swampman-Dolly would lack phenomenally conscious experiences.

Etiological representationalists can appeal to the fact that he has been copied from Davidson for holding that his mental states have phenomenal properties. Contrary to Swampman, who is not a copy of Davidson but the product of mere randomness, Swampman-Dolly inherits Davidson's historical properties. His mental states have functions in virtue of being a copy of Davidson. I fail to understand how copying could play the desired role here, but let me grant the adequacy of the reply and continue with the mental thought experiment.

Being able to produce relevantly similar creatures to us who lack phenomenal consciousness is a tremendously interesting project. Many would agree with the idea that if a zombie creature were to lack phenomenal consciousness then all kind of experiments on her should be allowed. Zombies do not feel any pain when the lancet cuts their skin or feel sad about the way scientists treat them. Investigation on zombies would surely lead to plenty of benefits for human kind. We would get the advantages of the investigation in humans avoiding most of the ethical reasons for not doing it. The project is nowadays feasible if

47 Totipotent stem cells can differentiate into embryonic and extraembryonic cell types. Such cells can construct a complete, viable, organism and are produced from the fusion of an egg and sperm cell.

etiological representationalism were true. Here is the recipe to produce them.

The Zombie Project

1. Take a random number generator that generates a sequence of 0s and 1s.
2. Use a computer to code pairs of numbers as: A (00), T (01), G (10) and C (11).
3. Connect the computer to a DNA synthesizer.[48] The DNA synthesizer receives the sequence from the computer and converts it into a molecule.
4. Group randomly these fragments of DNA and introduce them into a cell with the basic required proteins. The introduced genome is completely random and lacks any history.

The vast majority of the resulting combinations won't give rise to organisms, others will give rise to an organism but they will be unable to survive. However, the process will also give rise to dinosaur-like organisms, orangutan-like organisms and human-like organisms. According to etiological representationalism, this human-like organisms are zombies: lacking any evolutionary history and not being the copy of a human they lack function and thereby phenomenal consciousness.

The costs of producing human-like organisms can be reduced making the project economically feasible by previously filtering, computationally, the DNA random chains and introducing into cells only human-like DNA. Hopefully not many people would support this project, not being able to swallow the pill that these human-like organisms lack phenomenal consciousness.

Tye (2002) maintains that no causal history is required for having a function in a teleological account of mental content:

> What matters to the phenomenal content of a given state of an individual X is not necessarily any aspect of the actual causal history of X.
>
> The causal connections that matter to phenomenal content, I suggest, are those that would obtain, were optimal or normal conditions operative. (ibid. p.64)

In the case of the Swampman, Tye considers well-functioning conditions. Such conditions are met when there is an appropriate match between the creature's behavior and what is tracked in his environment. If her needs are fulfilled and flourishes then her states have content.

Tye seems to be endorsing an ad-hoc mixed position between etiological theories (for humans) and non-etiological accounts (for exotica).

> It might now be suggested that what the representationalist needs is a "mixed" theory of tracking in normal or optimal conditions. For creatures or devices with states that were designed to track things, for example, human beings and thermometers, those states acquire representational content at least partly via what they track under design conditions.

48 This is not science fiction but current state of the art.

> Here, if design conditions fail to obtain, then the setting is abnormal, no matter how long it obtains.
>
> For accidental replicas (for example, Swampman) the requirements are different. [...] Moreover, there are conditions under which he will flourish, and there are conditions under which he will not.
>
> This leads to the thought that Swampman can have inner states that acquire representational content via the tracking or causal covariation that takes place under conditions of *well functioning*. [...]Where there is a design, normal conditions are those in which the creature or system was designed to operate. Where there is no design, normal conditions are, more broadly, those in which the creature or system happens to be located or settled, if it is functioning well (for a sufficient period of time) in that environment.(ibid. p.121-122)

Block (2007b) has shown that Tye's strategy does not solve the representationalist's problems. The environment in Earth and inverted Earth matches equally well Swampman's behavior.[49] If swamp-grandchild travels to inverted Earth his behavior there is also well-functioning.

> So on what basis could Tye choose to ascribe to the swamp-grandchild the phenomenal character that goes with representing the Inverted Earth sky as blue (as a normal Earthian emigrant, according to Tye) rather than the phenomenal character that goes with representing the sky as yellow (like normal Inverted-Earthians)? A choice here would be arbitrary. Suppose Tye chooses the Earthian phenomenal character. But what makes that the privileged phenomenal character for the swamp-grandchild? The fact that his grandparents materialized on Earth as opposed to Inverted Earth? But that is a poor reason. Suppose the swamp-grandchild is born on Inverted Earth while his parents are on a visit and stays there. Are his phenomenal characters determined by his birth place or by his grandparents' birth place? There is no good reason for either choice and there is no plausibility in the idea that there is no matter of fact about what the phenomenal characters are. (ibid. p. 606)

One could resist Block's objection by insisting on the metaphysical impossibility of inverted spectrum scenarios. But Block's argument can be reproduced without any need of inverted spectrum scenarios. This will be considered in our next story, where I will make use of another variation of a famous mental thought experiment to create an additional argument against etiological representationalism.

A new kimu's tale

Etiological representationalism is committed to the metaphysical possibility[50] of microphysical duplicates that differ phenomenologically. All that is required is a different evolutionary history. That sounds very counterintuitive to me and gives rise to weird conclusions.

49 Tye accepts the metaphysical possibility of inverted spectrum.

50 Type-B materialists concede that microphysical duplicates that differ phenomenologically are conceivable, but they can resist their metaphysical possibility.

In Pietroski's imaginary tale (Pietroski (1992)) we are introduced to the kimus, simple-minded, colorblind creatures. Jack, a kimu, is born with a mutation, he has a new mechanism that produces a mental state, M, in response to red objects. M further causes Jack to approach red objects.[51] That leads Jack to climb to the top of the nearest hill every morning to see the rising sun. Luckily for Jack, he avoids the kimu's predators, the snorfs, who hunt in the valley below. The mental state M is inherited by Jack's descendant and the trait is selected for indicating something like *snorf free area* because *snorf free areas* and not *red objects* explain that M has been selected for.[52] According to etiological theories, M represents *snorf free area* and not *red objects*. Pietroski's intention is to press on the counterintuitive conclusion that a mental state with phenomenal character PC_{RED} does not represent red objects. Teleosemanticists, like Millikan (2000, p. 149), bite the bullet and accept this conclusion, distinguishing the properties that cause the representation from the content of the representation.

Both, Pietroski and Millikan, consider phenomenal properties to be intrinsic properties independent of the teleological content. My purpose now is to show the problems derived from making phenomenal properties identical to etiological representational properties; i.e. from accepting that subjects that differ in their causal history differ phenomenologically. For this purpose I will present my own variation of the kimu's tale.

Many years before Jack was born, due to tectonic movements, the kimus population was split into two groups. In the second group, Nuca is born at the same time as Jack with the very same mutation. There are no snorfs in the area inhabited by the second group. But on the top of the hill there is plenty of food. As a consequence of that, the trait, which is inherited by her descendants, is selected for indicating something like *food area*.

According to etiological representationalism, Jack and Nuca differ phenomenologically.[53] I hope some readers have already found this conclusion unacceptable, but let me continue with the story for those who remain skeptical.

Some individuals from Nuca's population migrate to an area with similar conditions to those in the area where Jack inhabited. Thanks to the old mutation they avoid snorfs. After several generations the content of their mental states changes to *snorf free area* or maybe to *snorf free area or food area*. The phenomenal content of their experience accordingly changes. According to etiological representationalism, for each of Nuca's successors there are only three possible candidates for the content: *snorf free area*, *food area* and *snorf free area or food area* and consequently there must be pairs of individuals A and B such that A is the direct ancestor of B and they differ phenomenologically. That seems to me a really hard pill to swallow, especially if we consider that, for all that matters, A and B could be microphysically identical.

51 In the original story it is granted that Jack enjoys a sensation as of red and what is questioned is the content of this state. I am making use of Pietroski's nice tale for a different purpose.

52 Tye, for instance, requires the content to be poised for reasoning and motor control. If kimus are too simple for having the required system, consider kimas instead. Kimas are as similar as possible to kimus but having the required system for the content to be poised in this way.

53 Within this framework, Block's argument against Tye's position can be presented: which would be the phenomenal character of swampkimu?

Alternatively, my oponent can appeal to vagueness. She can claim that it is indeterminate what A and B represent and that the phenomenal characters of their respective experiences are borderline cases. They neither determinately indicate *snorf free area* nor determinately indicate *food area.*

To illustrate the problems of such a reply and to show that we do not need to appeal to weird creatures to find etiological representationalism problematic I am going to present my last original objection to etiological representationalism.

Vertical vagueness

In order to have certain content C, M has to have been selected for. It is not only the causal role of the state, but also its history that explains the content that M has. According to teleosemantic theories, when M appears for the first time, it lacks content.[54]

We want to know what a qualitative property is; i.e. we want to know which is the property that determines the concrete phenomenal character of the experience, for instance of an experience with phenomenal character PC_{RED}. Etiological representationalism tells us that it is the property of representing a certain content, say C_{RED}.[55]

Assume that M appears in a certain individual S_1 whenever she is in front of a certain property C_{RED}. In a modern subject, S_n, who is a descendant of S_1, the mental state M represents C_{RED} because indicating C_{RED} is what explains M being there. According to etiological representationalism, in S_1 M does not represent anything, therefore, according to etiological representationalism there is nothing it is like for S_1 to be in M. She doesn't feel anything; there is no representational content, so there is no phenomenal character. On the other hand, S_n, a modern subject, has an experience with qualitative character PC_{RED} because M represents C_{RED}. A subject has an experience with a certain qualitative character because she is in M and M represents certain content. Neither S_1 nor S_2 (the direct descendant of S_1) nor S_3 (the direct descendant of S_2) instantiates a mental state that represents C_{RED}. On the opposite side, S_n, S_{n-1}(the direct ancestor of S_n), S_{n-2} (the direct ancestor of S_{n-1}) can be in a mental state that represents C_{RED} and therefore they feel something while looking at C_{RED} objects.[56]
There is a range of individuals for which it is indeterminate whether or not they instantiate C_{RED}. We cannot ascribe them with the corresponding content, we cannot ascribe them with any content. The situation

54 Block (2007b) suggests that etiological representationalism is committed to accept that swampchildren inherit the lack of content.

> ...since phenomenal character is a kind of representational content that derives from evolution, then swampchildren have no phenomenal character. Zombiehood is hereditary (ibid. p.603)

This cannot be right, for it would make etiological representationalism implausible. When a mutation appears for the first time it lacks content; evolution is required for intentional content. When the mutation is selected for, several generations later it will be a contentful state.

55 C_{RED} is an appearance feature, A_{RED}, if my previous analysis is right. The argument I am going to develop against etiological representationalism is independent of the previous analysis, so I will refer to the content as C_{RED}.

56 It is very important to note that all of the individuals in the series can be functionally identical (the notion of function here is the classical one where functions are understood as causal roles) with regard to whatever that plays a role in content selection and in any other respect.

can be summed up as follows:[57]

$$S_i \begin{cases} \text{does not instantiate } PC_{RED} & i \leqslant l \\ \text{indeterminate whether instantiates } PC_{RED} & l \lesssim i \lesssim m \\ \text{instantiates } PC_{RED} & i \geqslant m \end{cases}$$

Etiological representationalism requires that there be borderline cases not between $PC_{RED_{34}}$ sensation and $PC_{RED_{33}}$ sensation, but rather between $PC_{RED_{34}}$ and no sensation at all. Etiological representationalism requires vertical vagueness.

In chapter 3 I defined the notion of vertical vagueness as follows:

VERTICAL-VAGUENESS The phenomenal character Q of an experience of a subject S is vertically vague if and only if it can be indeterminate whether S's experience has phenomenal character Q or no phenomenal character at all.

This notion will be useful for the objection I am going to present against etiological representationalism. Vertical vagueness deals not with what you feel, but with whether you feel. The fact that phenomenal characters are not vertically vague seems to me to be a truism. Having a concrete experience, instantiating a concrete phenomenal property, is a matter of on/off. For a subject S, either she feels something or she doesn't, either there is something it is like to be in the mental state she is in or there isn't. We cannot make sense of a borderline case between a horrible headache and no pain at all.

The problem is that *having content A* is a vague property, according to etiological representationalism, and so should be the property of *having an experience with phenomenal character* PC_A. [58] But having *having an experience with phenomenal character* PC_A cannot be vague in the vertical sense, the sense required by etiological representationalism. So, phenomenal properties cannot be representational properties if representational properties are to be understood as teleosemantic theories maintain.

The only way I can conceive of a sensation fading over a series is through other, different sensations. Imagine you are feeling a horrible headache. You take a painkiller and concentrate in the pain you are feeling. If you were asked after half an hour whether your sensation is the same as when you had to take the painkiller, you would surely reply that it isn't. After an hour you have no pain at all. The pain has gone through a series of states, each of them FS-indistinguishable from the previous and subsequent. Representationalists would explain this difference in character as a difference in the content. Such an explanation is not available for the series of individuals along the selection process, because the content, if any, is the same in all cases.

When discussing about fading qualia, Chalmers (1996) considers the possibility of vertical vagueness in the case of Joe, whose neurons are replaced by silicon chips, as we saw in 2.1.2.

57 In the example above, I am considering natural selection as the selection process for M, where several generations are required for the selection of the trait and, therefore, for the trait to have a teleological function. Some teleosemantic theories defend other selection processes. The objection applies also *mutatis mutandis* to these theories.

58 I am assuming here that the V-identity principle is right. The V-identity principle, introduced in chapter 3, maintains that two properties cannot be identical unless they share their borderline profiles.

> Let us focus in particular on the bright red and yellow experiences I am having from watching the players' uniforms... Perhaps he [a subject with faded experience -Joe] is having the faintest of red and yellow experiences. Perhaps his experiences have darkened almost to black. There are various conceivable ways in which red experiences might gradually transmute to no experience at all, and probably even more ways that we cannot conceive. But presumably in each of these the experiences must stop being bright before they vanish...imagine that Joe sees a faded pink where I see bright red, with many distinctions between shades of my experience no longer present in shades of his experience. (ibid. pp. 238-239)

A bright red experience and a faint red experience are FS-distinguishable, and so are red and faded pink. These experiences do not have the same phenomenal character and, according to representationalism, they have different content. However, for all the individuals in the evolutionary chain there is a single possible content and therefore one single phenomenal character.

In a nutshell, the problem is that etiological representationalism requires all phenomenal properties to be vertically vague, but they cannot be, so etiological representationalism is wrong.

Let me now consider some possible replies that the etiological representationalist might make and try to show that they cannot succeed.

POSSIBLE REPLIES

Reply 1: *Dull* Experiences In the previous chapter I considered the case of dull experiences as a candidate for vertically vague experiences. One could try to resist the intuition that supports my argument along these lines. Let me repeat Tye (1996)'s quote :

> [C]an it be vague whether a given state is an experience, whether there is anything at all it is like to undergo the state? It seems to me that it is not pre-theoretically obvious that the answer to this question is 'No'. Suppose you are participating in a psychological experiment an you are listening to quieter and quieter sound through some head-phones. As the process continues, there may come a point at which you are unsure whether you hear anything at all. [...]it could be that there is no fact of the matter about whether there is anything it is like for you to be in the state you are in at that time. In short, it could be that you are undergoing a borderline experience. (ibid. p.682)

According to Tye, there is no pre-theoretically obvious answer to the question as to whether or not a phenomenal character is vague. According to etiological representationalism, having a certain content is a vague property, and so are phenomenal properties.

Rejoinder This line of objection is based on a misunderstanding of the intuition that supports my argument. I was pointing out that we have a clear pre-theoretical intuition that, at the very least, certain types of sensation cannot be vertically vague.

In the examples presented by Tye, the assumed vagueness is explained as a variation in the intentional content. One can concede to Tye, for the sake of discussion, that in the example he presents, a case of a *dull* experience, it is vague whether or not the subject feels one thing or other, or even that it can be vague whether the subject feels anything at all. Etiological representationalism requires, however, that the phenomenal character of *all* experiences is vertically vague. This possibility is not available in the case of a vivid sensation, like the horrible headache I am having right now or the experience I am having while looking to the red apple.

For all the individuals in the evolutionary chain the content, if there is one, is always the same: if their mental state is about something, it is about C_{RED}. The point is that, according to etiological representationalism, since the only possible content for every subject in the chain (same internal state and same environment) is C_{RED}, either she feels the same that a modern individual feels, or she doesn't feel anything at all.

One could claim that there is a sense in which, for a subject S, it may be indeterminate whether or not she feels something. We cannot ascribe the phenomenal property to S because it is indeterminate whether or not she represents the corresponding content. S will, however, either feel PC_{RED} or she won't. Although one can make sense of borderline cases between a very light noise and no sound at all, one cannot make sense of a borderline case between PC_{RED} and no sensation at all or between a horrible headache and no pain at all.

Reply 2: Epistemicism or Radical Emergentism. Recall that, as we saw in the previous chapter, epistemicism (Sorensen 2001; Williamson 1996) claims that vague predicates have sharp borders but we are cognitively closed to them.

In this particular case, the advocate of vagueness as an epistemic problem would maintain that we cannot set the individual from which the proper content can be ascribed. There is nevertheless a precise value of i between l and m, though we cannot know it.
Imagine that j is the value such that:
$\forall i(i \lesssim j \rightarrow S_i \, does\, not\, instantiate\, PC_{RED})$
$\forall i(i \gtrsim j \rightarrow S_i \, instantiates\, PC_{RED})$

Rejoinder It seems to me that selection is essentially a matter of degree. If this is true, it is hard to believe that there is a precise instant i where the function has been selected, such that for every instant before i the function was not selected.

Be that as it may, acceptance of epistemicism leads to the following unacceptable conclusion: S_j feels something while looking at a red object, whereas her immediate ancestor (S_{j-1}) does not feel anything. The only difference between S_{j-1} and S_j is one generation. Nevertheless, S_j's mental state has a content that S_{j-1} lacks and there is no further physical difference between the two subjects. It is hard to believe that S_j feels something while S_{j-1} feels nothing at all. This solution doesn't seem plausible.

In conclusion, accounts of function based on evolution seem to be unsatisfactory for our purposes.[59] If we want to grant phenomenally conscious experiences to creatures that lack evolutionary history (Swampman) and avoid the former objections, then qualitative properties cannot be etiological functional properties.

Etiological theories of mental content are not suitable candidates for explaining the relation of representation in the case of phenomenally conscious experiences if qualitative properties are to be identified with representational properties. If any functional theory is to succeed for accounting for phenomenal properties, we should better have a satisfactory notion of function that does not depend on the trait's history. Non-etiological functions are an attempt in this direction.

4.4.2 *Non-Etiological Functions*

We are looking for a theory of functions that allows us to unpack the normal conditions in (Self-attributed) and explain in naturalistic terms the relation of representation. We have seen that etiological theories are not a suitable candidate.

It is controversial that this is the right analysis of the notion of function in Biology. Cummins (1975) emphasizes that the explanatory role of function attributions is to explain a capacity to which the exercise of the function contributes, rather than to explain the presence of the function bearer which is precisely what etiological accounts do.[60] In contrast, other philosophers maintain that their definition correctly describes certain uses of the term in Science.

Non-etiological theories of function claim that the function of a trait is related to what the trait actually does. Ruth Millikan (1984), who has developed the most thoughtful account of teleosemantics, agrees (Millikan, 1989), but claims that there is another sense of function, as selected effect, that intrinsically explains the presence of the trait. She has suggested that this notion of function is an stipulated definition that has to be judged by its utility in solving philosophical problems, independently of whether or not this is the notion used in fact in any scientific field such as Biology.[61]

Most philosophers seem to agree that etiological and non-etiological notions are complementary notions (Godfrey-Smith (1994); Millikan (1989)). The explanation of a functional trait seems to be conceptually independent from the explanation of its contribution to a capacity of the system. Non-etiological functions search for a criterion for selecting one among the several causal roles of the trait. From all the trait does, these theories select one (or more) of them as relevant, as the function of the trait.

---

59 And *mutatis mutandis* any etiological theory of function. Note that the problems derive just from the requirement of a causal history, not from the fact that such a history is evolutionary.

60 See Wouters (2005) and Boorse (2002) for a convincing rejection of the claim that functions in biology are something a trait is supposed to do as opposed to what it does. According to Wouters, a view that maintains that 'performing a function' is parasitic on 'having a function' puts the cart before the horse. They suggest that claiming that a trait has a function it does not perform is just another way of claiming that the trait does not perform the function an homologous item performs. There are exceptions to generalizations but those exceptions are not something specific to functions.

61 Millikan has expressed, in conversation, her rejection of teleological theories of mental content as a plausible account of phenomenal characters.

There is no single and unified non-etiological definition of function. Non-etiological theories of functions can be divided into two groups: systemic and goal-contribution.

Systemic theories are inspired by Cummins (1975)'s analysis of function as the causal contribution a structure makes to the overall operation of the system that includes it. Cummins' concept of function is not a historical or evolutionary concept. According to Cummins, a component may have a function even if it was not designed or selected for and, therefore, parts with no selection history can be ascribed a function. The function of a trait is its contribution to the system it belongs to.

Goal-oriented theories are classically inspired by Nagel (1961). According to these theories, the function of a trait is its contribution to the organism's goal of survival and reproduction. To a first approximation, we can say that a system has a goal G if (within certain boundary conditions) it is disposed to vary its behaviors in the manner required to achieve or maintain G.

The main disagreement between systemic and goal-oriented theories is on whether function attributions depend on the way the system behaves (goal-oriented) or the way it is organized (systemic).

Both theories face serious problems for explaining the content of our mental states. We are looking for a theory of function that allows us to tell apart cases in which the trait is functioning correctly from cases in which it is not. In particular, on the assumption that mental states are states of the nervous system, we are looking for cases that tell us which is the function of certain nervous states. Consider a nervous state N_{RED} which is the neural correlate of experiences with phenomenal character PC_{RED}. Whenever N_{RED} is activated, the subject undergoes an experience with phenomenal character PC_{RED}. N_{RED} is activated both when I am confronted with red apples and when I take the LSD. Both, being activated by A_{RED} things and being activated by $PILL_{RED}$ things is part of its causal role. Both, A_{RED} things and $PILL_{RED}$ things have the disposition to cause experiences with phenomenal character PC_{RED}, but my experience is about A_{RED} and not about $PILL_{RED}$ because A_{RED} and not $PILL_{RED}$ is what causes the experience *in normal circumstances*. In this case, our desired theory of function should tell us why indicating A_{RED} and not indicating $PILL_{RED}$ is the function of N_{RED}; why being caused via the visual path and not via the vasculatory system *is normal*; i.e. why when N_{RED} is activated via the vascular system it malfunctions.

Systemic theories can hardly account for these differences, for N_{RED} makes both contributions to the system: being activated when the corresponding input comes from the visual path and being activated when the corresponding input comes from the blood flow. These are two things that N_{RED} does. How can we decide among them which is the function of N_{RED}?

Goal-oriented theories fare a bit better, but are still unsatisfactory. For goal-oriented theories the function of a trait depends on the goal of the system. Here, they can make a principled distinction between A_{RED} and $PILL_{RED}$: indicating A_{RED} and not indicating $PILL_{RED}$ contributes to the goal of the system. Unfortunately, I see two main problems for goal-oriented theories.

The first one is to provide an account of the goal of a system. This notion seems equally normative and we would be just shifting the

problem of normativity from the function of the trait to the function (goal) of the system.

The second one is that even if we were able to provide a materialistic justification of which is the goal of the system, anything that would contribute to this goal would be a function. Imagine that the goal of beings like us were something like surviving and reproducing. N_{RED} has the function of indicating A_{RED} because indicating A_{RED} contributes to this goal. The problem in this case is that any other causal role that N_{RED} has, like indicating $PILL_{RED}$, would also be a function if it also contributes to this goal even in weird conditions.[62]

It seems to me that the kind of norms required for a naturalistic theory of the mental content of thoughts, beliefs or desires can hardly be provided by a non-etiological theory. I do not see how we could explain why the mental state of the frog is about flies and not about black specks, to take the famous example, without appealing to the environment in which the neural state was selected for.

In spite of this, it is not clear to me that further developed non-etiological accounts cannot provide the kind of norms required for the content of phenomenally conscious mental states. In the case of phenomenally conscious mental states we only need to discriminate among different ways of causing the experience. Red apples, LSD and mad scientists manipulating the brain can activate N_{RED}, what is required is, I think, a naturalistic way of discriminating one way of causing the experience as the *right* one, objects that can cause the experience that way will be A_{RED} objects. However, something more is required for naturalizing our semantics because both black specks and flies can cause the state in very same way.

Merely as an illustration of what these theories could look like I will present organizational accounts.[63]

Organizational Accounts

Non-etiological accounts maintain that the function of a trait is the trait's contribution to the system. Millikan (2002) has objected that there seems to be no independent way of choosing among the many systems (the visual system, the nervous system, the organism, etc). If we want non-etiological accounts to get off the ground we need an objective way to determine the system of reference.

Mossio et al. (2009) have presented a non-etiological theory known as Organizational Account where they appeal to the notion of *self-maintaining system* to provide such a reference. Biological systems are sophisticated and highly complex examples of natural self-maintaining systems.

In a self-maintaining system the dynamics of the system tends to maintain the inherent order; its organizational pattern appears without a central authority or external element imposing it through planning. This globally coherent pattern appears from the local interaction of the elements that make up the system. The organization is in a way parallel,

62 Restricting such conditions to normal circumstances would obviously not be of any help, for it is precisely the notion of normal circumstances the one we want to unpack.

63 Alternatively, (Schroeder (2004) offers a cybernetic account in which the systemic account is supplemented with a cybernetic system driving some neural mechanism toward producing tokens with the relevant causal role R. If some inner system monitors the production of states with this causal role R, and "rewards" it for success while "punishing" it for failures then the state will have R as its function. It is an empirical open question whether such a cybernetic mechanism exists or not.

for all the elements *act* at the same time and is also distributed for no element is a coordinator.

The notion of self-maintained system has a long history in philosophy dating back to Aristotle (Godfrey-Smith (1994); McLaughlin (2001)). In contemporary science it was popularized by cyberneticians. More recently, after Ilya Prigogine won the Nobel Prize in 1977 for his work on dissipative structures and their role in thermodynamics, many scientists start to migrate from the cybernetic approach to the thermodinamic view on self-maintaining systems.

The minimal expressions of self-maintenance are 'dissipative structures':

> Dissipative structures are systems in which a huge number of microscopic elements adopt a global, macroscopic ordered pattern (a 'structure') in the presence of a specific flow of energy and matter in far-from-thermodynamic equilibrium (FFE) conditions. Mossio et al. (2009, p. 822)

A simple example of these self-maintained systems is the flame of a candle. In the flame of a candle, the microscopic reactions of combustion give rise to a macroscopic pattern, the flame, which makes a crucial contribution to maintain the microscopic chemical reaction by vaporizing wax, keeping the temperature above the combustion threshold, etc. The flame itself favors the conditions that enable it to work.

Self-maintaining systems are *organizationally closed*. There is a circular causal relation between some higher level pattern or structure and the microscopic dynamics and reactions, as the candle example illustrates. The organizational closure provides a criterion for the goals of the system: in an organizationally closed system the goal states are the stability points through which the system can exist (Barandiaran and Moreno (2008)). Mossio et al. understand organizational closure in such a way that "the activity of the system becomes necessary (even if, of course, not sufficient) condition for the system itself." Mossio et al. (2009, p. 824)

The last notion required for the Organizational Account is that of *Organizational differentiation*. A system is organizationally differentiated when it is possible to distinguish parts that contribute in different ways to the self-maintenance of the system. More precisely:

> Organizational differentiation implies not only that different material components are recruited and constrained to contribute to self-maintenance but, in addition, also that the system itself generates distinct structures contributing in a different way to self-maintenance. In other words, material components become candidates for functional attributions only if they have been generated, and are maintained, within and by the organization of the system. A self-maintaining system is organizationally differentiated if it produces different and localizable patterns or structures, each making a specific contribution to the conditions of existence of the whole organization.Mossio et al. (2009, p. 826)

With these tools in hand, we can present the Organizational Account of function as:[64]

64 Mossio et al. definition of the function of a trait according to the theory is incomplete as Artiga notes. My presentation departs slightly from both. The differences are irrelevant

> (Organizational Account)
>
> A trait T has a function F if and only if:
>
> (C1) T's performance of F contributes to the maintenance of the organization O of S.
>
> (C2) T is produced and maintained under some constraints exerted by O.
>
> (C3) O is organizationally closed.
>
> (C4) S is organizationally differentiated and T is one of the parts in which S is differentiated.

According to the definition, the function of the heart is pumping blood since pumping blood contributes to the maintenance of the organism by allowing blood to circulate, which in turn enables the transport of nutrients to and waste away from cells, etc. Additionally, the heart is produced and maintained by the organism, whose integrity is required by the heart itself. Furthermore, the system is functionally differentiated: it produces other structures that contribute in different ways to the maintenance of the system.

There are plenty of functional traits in systems like us that are functional without being necessary for the system's existence. We want our theory to be able to attribute a function also to these traits. For that purpose (C1) should be refined:

> (C1′) T's performance of F contributes to the maintenance of O in the sense that that specific organization would not exist without T.

(C1′) attributes a function to a trait even if the presence of the trait is not necessary for the self-mantaining system. F is the function of T because if T would not do F either S would thereby cease to exist or it would continue existing with a different regime of self-maintenance, where *regime of self-maintenance*[65] is to be understood as possible specific organizations of the subject S.

Mossio et al. illustrate this by considering two typical biological functions in humans: the heart's pumping blood and the eyes' transduction of light. In the first case there is no possible organizational alternative and if the heart does not pump blood the self-maintained system ceases to exist. In the case of the eye there is an alternative regime of self-maintenance.

> In the first case, the functional trait is indispensable, because it contributes to generating a global process (the circulation of blood), which is required in order to preserve the existence of this class of systems, whatever regime of self-maintenance of the members is considered. In this sense,

for my purposes which are merely illustrative of a possible non-etiological account of function.

65 Mossio et al. define regime of self-maintenance as follows:

> We call regime of self-maintenance each possible specific organization that an individual member of a class can adopt without ceasing to exist or losing its membership of that class. Each class may thus include several regimes of self-maintenance. In organizational terms, if a trait is subject to closure (and thus has a function), then the specific regime of self-maintenance that the system has adopted requires the said trait as an indispensable component. (Mossio et al., 2009, p. 829)

> there are no organizational alternatives to blood pumping for humans to be viable. In the second case, in contrast, the transduction of light contributes to generating a capacity (seeing), which constrains other processes in specific modes of self-maintenance but is not indispensable for human beings (blind people can survive). However, since the transduction of light is functional, this crucially means that a whole network of processes depends in some way on the capacity of the eyes to transduce light. Accordingly, if the eyes were to stop performing their function or if they were to malfunction, a global constraint (vision, in this case) would disappear and the system would be forced to shift to a new regime of self-maintenance (in this case, find new ways of finding food, moving around; etc.) (ibid. p.830)

Organizational accounts explain cases of malfunction when a trait satisfies C2, C3 and C4 but fails to satisfy C1. If the eye activates the corresponding neurons without having being stimulated by the corresponding wave-length it is malfunctioning, because the eye satisfies C2, C3 and C4, but not C1, it is not transducing light.

Let's concentrate now on the case of phenomenal properties. How does the organizational account provide a content for phenomenally conscious mental states?

The function of N_{RED} is to indicate A_{RED} and not to indicate $PILL_{RED}$ because N_{RED} being active when there is an stimulus coming from the visual path contributes to the organization of the self-maintained system and being active when the stimulus is coming from the vascular system does not. In both cases (C2), (C3) and (C4) are satisfied. It is an open empirical question whether the former statement is true.

Much more work has to be done for better clarification of the notions involved as those of *regime of self-maintenance, organizational differentiation,* etc in order to properly evaluate the merits of the proposal. My purpose by presenting organizational accounts was simply to illustrate how a non-etiological theory could account for the relation that holds between the representation and the representata in the case of phenomenally conscious mental states.

4.5 THE QUALITATIVE CHARACTER OF EXPERIENCE

To close this chapter let me recapitulate and properly elaborate on basic aspects of the theory that I am propounding.

The phenomenal character of an experience E of a subject S is the way it is like for S to undergo E. When I look at the red apple there is a *redness way it is like for me* to see the apple. This *redness way it is like for me* to see the apple is the phenomenal character of the experience. I have been referring to this phenomenal character as PC_{RED}.

Phenomenally conscious experiences are a certain kind of mental states: phenomenally conscious mental states. These mental states have a property that non-phenomenally conscious mental states lack: there is something it is like for me to be in one of the former states and nothing in one of the latter ones.

Furthermore, we distinguish different kind of experiences. Compare the experience you have while looking at a red apple with regard to a particular property, the color for instance, with the experience you have

while looking at the putting green in a golf course. There is a *redness* way it is like for you to have the former experience and a *greenness* way to have the latter. In these situations, you would be undergoing two different mental states, M_1 and M_2. These mental states have different qualitative properties. Qualitative properties determine the differences in phenomenal character of experiences; they determine the concrete phenomenal character that the experience has.

Whenever I am in a mental state I am in a distinctive brain state. When I undergo an (token of) experience with phenomenal character PC_{RED} I am in a (token of) a brain state N_{RED}. N_{RED} is the neural correlate of my experience with phenomenal character PC_{RED}, the neural activity that in *my case* perfectly correlates with experiences with phenomenal character PC_{RED}.

Some type identity theories hold that my property of having an experience with phenomenal character PC_{RED} is identical to my property of having N_{RED}. According to these theories, phenomenally conscious states are identical to brain states. The problem of this proposal is that we don't want to maintain that all the properties that N_{RED} has are necessary for having an experience with phenomenal character PC_{RED}. We don't even want to claim that the essential properties of N_{RED} are necessary for having such an experience. If the intuition that replacing neurons by silicon chips (functionally identical to neurons) does not affect phenomenal consciousness is right, then we can replace my neurons with silicon chips to obtain $Silicon N_{RED}$. In this case, I would not have N_{RED}, but I would undergo a phenomenally conscious experience with phenomenal character PC_{RED}. The property of having an experience with phenomenal character PC_{RED} cannot therefore be identical to the property of having N_{RED}. Of course, N_{RED} and $Silicon N_{RED}$ share some properties; the question is which of these properties are necessary for phenomenal consciousness. Representationalism can offer a partial reply to this question.

Representationalism offers an account of qualitative properties. According to representationalism, qualitative properties are representational properties. Differences in the phenomenal character of the experience are differences in the content of the experience. Qualitative properties are the properties in virtue of which the state satisfies a certain functional role: the function of indicating a certain feature. N_{RED} is one of the possible realizers of this functional role.

I have been arguing in favor of a particular version of narrow representationalism. Narrow representationalism holds that the representational properties that account for the differences in the phenomenal character are intrinsic properties of the individual: microphysical duplicates undergo the very same experiences. To make compatible the internalist intuition with representationalism and to avoid defeating problems, I have appealed, following Shoemaker and Egan, to appearance properties: centered features of being disposed to cause the experience.

To say that the content of the experience is a centered feature is to say that it is a function from pairs of possible worlds and individuals to extensions. Consider the following centered feature: being disposed to cause experiences with phenomenal character PC_{RED} in me. When we introduce the actual world and Sebas as arguments of this function we obtain an extension that includes among other things red apples, holograms and LSD.

We do not want, however, to say that my experience represents anything that is disposed to cause the experience: I do not attribute any feature to the LSD when I undergo a PC_{RED} experience.[66] For that reason we appeal to normal circumstances: red apples and holograms are disposed to cause an experience with phenomenal character PC_{RED} in me in normal circumstances.

The relation of representation is normative; in order to unpack this notion we appeal to the notion of function. In the previous section I have tried to show the problems derived from appealing to an etiological theory of functions. If these arguments are sound, then phenomenal properties cannot be determined by any etiological function because we don't want phenomenal properties to be dependent on the causal history of the state. If no non-etiological account of function could explain the representation relation, qualitative properties would not be representational properties because the differences in phenomenal character would not be differences in the intentional content.[67] Swampman would then lack any intentional content (if the relation of representation can only be naturalized via etiological functions) and, nevertheless, the experiences it would have while looking at a red apple and when looking at the grass would differ in character. Furthermore, it is hard to make sense of the idea that Swapman can undergo the very same phenomenally conscious experiences as I undergo, while maintaining that its experiences do not represent anything.

I have expressed, and given reasons for, my confidence in a non-etiological theory of function that satisfactorily explains the content of phenomenally conscious experiences. I will appeal to the organizational account presented in the previous section to illustrate my examples.

I claim that the properties of my brain state in virtue of which my experience has the phenomenal character it has (qualitative properties), PC_{RED}, are a subset of properties of N_{RED}: the properties that are necessary for satisfying a certain functional role. Qualitative properties are the properties in virtue of which my token of N_{RED} *represents* A_{RED}.

I cannot offer a full picture of the view at the moment. I will do it in the next chapter (5.4.2 and 5.4.3). We can, nevertheless, make some clarifications. For that purpose, I will consider a theory of appearance properties in the vicinity of the one I have made. I will consider that the content of an experience with phenomenal character PC_{RED} is the *property* of being disposed to cause an experience with phenomenal character PC_{RED} in me (or in Sebas-type individuals). This proposal corresponds to what I have called (Indexical disposition de re), as we have seen on page 138. I have argued, following Egan (2006a), that this cannot be the theory we are looking for, because the content of the experience, in this case, is a property and not a centered feature. (Indexical disposition de re), contrary to (self-attributted*), fails to

66 In other words, in having an experience with phenomenal character PC_{RED} I do not self-ascribe the property of being presented with any of the properties of the LSD.

67 As stated by non-etiological theories of function, a trait has a function if and only if it satisfies a certain causal role within the system. According to etiological theories, besides a causal role an evolutionary history is also required for the trait to have a function.
I think that the best shot for those who want to hold etiological theories of content is to accept that Swampman has phenomenally conscious experiences, deny that qualitative properties are representational properties and identify qualitative properties with the properties in virtue of which the state *comes to have the function of indicating;* namely, the properties in virtue of which, if Swampman were to have the same evolutionary history as I have, it would represent the very same features that I represent.

satisfy POS-SAMENESS: an individual that is not of Sebas-type[68] cannot undergo experiences with the same phenomenal character as I do. Furthermore it fails to satisfy INCOMPATIBILITY (see page 140). Nevertheless, this proposal will be useful to clarify the function of a phenomenally conscious mental state and to dispel the worries about circularity.

Let's assume that N_{RED} is the neural correlate of my experience with phenomenal character PC_{RED} (the neural activity that in my case is minimally sufficient for having an experience with phenomenal character PC_{RED}) and that the organizational account can naturalize the relation of representation.

Let O be my organism. O is a self-maintaining system. Let me assume that:

> (C2)N_{RED} is produced under some constraints exerted by O.
>
> (C3) O is organizationally closed.
>
> (C4) S is organizationally differentiated and N_{RED} is one of the parts in which S is differentiated.
>
> In this case the function of N_{RED} is to indicate a certain property, which we can call P_{RED}, if and only if:
>
> (C1′) N_{RED} indicating P_{RED} things contributes to the maintenance of O in the sense that its specific organization would not exist without N_{RED}.

N_{RED} indicates anything that can produce it. However, indicating P_{RED} things contributes to the maintenance of O. Therefore, the function of N_{RED} is to indicate P_{RED} things. There are other things that N_{RED} indicates, other things that can produce N_{RED}, but N_{RED} has the function of indicating P_{RED} things, because N_{RED} contributes to the maintenance of O by indicating P_{RED} things. Assuming that a trait represents P_{RED} if, and only if, the trait has the function of indicating P_{RED}, then N_{RED} represents P_{RED}.

What is relevant for having the function of indicating P_{RED} things is that N_{RED} satisfies a certain causal role. Any state that satisfies this causal role in my organism would be a state that has the function of indicating P_{RED} things. Qualitative properties are the properties that are necessary and sufficient for satisfying such a causal role. If a silicon network ($SiliconN_{RED}$) can satisfy this causal role then we could replace N_{RED} by this silicon neural network and I would still undergo the very same kind of experience when $SiliconN_{RED}$ is activated.

We lack enough knowledge about the organism and the theory of functions that I have presented is still not suitably developed to provide a complete characterization of the function of N_{RED} and therefore of the corresponding properties. I am going to use a simplistic characterization of this function. Let's assume that the function of N_{RED} is to indicate those things that can produce it via a particular visual path under certain lighting conditions. In this case it is plausible that the causal role required would include properties as being connected to the appropriate eye cells, for instance.

68 The individuations of types of individuals depends on a theory of function. For example, according to the organizational account two individuals are of the same type if and only if they are self-mantaining organisms with the same organization. Other theories may set other constraints; for instance, an etiological theory would demand that individuals share a significant evolutionary history for them to belong to the same type.

I have been talking about the content of an experience with phenomenal character PC_{RED} as the centered feature of having the disposition to cause the experience in normal circumstances. This is clearly circular. On the one hand, we characterize the content of the experience as the centered feature of being disposed to cause the experience in normal circumstances and on the other hand, we individuate the experience by its phenomenal character; i.e. by its representational content.

There is, I think, only an apparent circularity in this proposal. Once all the terms are clarified, on the plausible assumption that necessary coextensive properties are identical, the circularity disappears. I will use (Indexical disposition de re) for the elucidation. The apparent circularity is the same one in (Indexical disposition de re) and in (Self-attributed*), one should merely rephrase the previous paragraph substituting *centered feature* by *property* to realize that the problem is the same.

(Indexical disposition de re) maintains that the property of *having an experience with phenomenal character* PC_{RED} is the property of *being in a state that has the function of indicating what can produce it via a particular visual path under certain lighting conditions* (P_{RED}). In my case, the realizer of this function is N_{RED}.

N_{RED} indicates many things, however it represents only those things that can cause the experience in normal circumstances, normal circumstances that are unpacked by the function of the state. If we can individuate the properties that my mental state represents without appealing to the state it can cause, then there will be nothing circular.

Something is P_{RED} only if it is disposed to cause N_{RED} in my organism in normal circumstances. These normal circumstances are unpacked by the function of N_{RED}. Hence, something is P_{RED} only if it is disposed to produce N_{RED} in my organism via certain visual path under certain lighting conditions. The apple, the hologram and the LSD are disposed to produce N_{RED} (N_{RED} indicates all of them) in my organism. However, the apple and the hologram but not the LSD are disposed to produce N_{RED} via the particular visual path under the appropriate lighting conditions.

In order to be able to produce N_{RED} in my organism via the appropriate visual path under the appropriate lighting conditions the object has to cause the excitation of certain eye's cells in this lightning conditions. To this effect, the object has to either reflect light (in this lighting conditions) or to emit light with a certain wavelength (627-770nm.). Therefore, something is P_{RED} if and only if it can emit or reflect light with certain wavelenght. According to (Indexical disposition de re), the property of having the disposition to cause an experience with phenomenal character PC_{RED} in Sebas is the disjunctive property of reflecting light (in this lighting conditions) or to emit light with a certain wavelength. This can be extended *mutatis mutandis* to other kinds of experiences like taste experiences, auditory experiences, etc. I think that there is nothing circular in this proposal.

The former proposal is only presented to dispel some worries about circularity. It cannot, however, be the proposal we are looking for. According to it, qualitative properties, the properties that my brain state has such that when I am in this state I undergo a phenomenally conscious experience with a concrete phenomenal character, are representational properties. However, in this example the content that determines the concrete kind of experience is a property. We have seen,

following Egan, that the content that determines the concrete kind of experience should be *de se*. The content is not a property but a centered feature. We will see the elaboration of (self-attributed*) at the end of the next chapter in section 5.4.

One consequence of the dispositionalist proposal in (self-attributed*) that some readers might find puzzling is the idea that centered features and not properties enter the content of the experience. Egan presented this proposal as a price worth paying for saving representationalism. In the next chapter I will argue that there is no withdrawal in this view and that a careful analysis reveals precisely that the content of the experience is de se. Thoughtful reflection and observation of the phenomenal character of our experiences reveals that in having an experience I attribute a certain property to myself and that when I introspect I do not find any property of my state but a self-attributed property.

There are some questions that remain unanswered:

- What do different phenomenally conscious states have in common?
- I have said that the causal role selected by (Indexical disposition de re) is not sufficient for having a phenomenally conscious experience. A mental state that satisfies this causal role will represent a property. I have maintained that the content that determines the concrete kind of experience the subject undergoes is *de se*. What is the causal role that phenomenally conscious states, and not other states, satisfy such that when someone is in this state she undergoes a concrete kind of experience and a phenomenally conscious experience at all?
- How do we come to self-represent a certain content?

These questions are the topic of the next chapter: the subjective character of experience.

5 THE SUBJECTIVE CHARACTER OF EXPERIENCE

This morning I went to a fruit shop. After some time my red apple was starting to rot and, as the reader has already noted, I cannot work without a red apple. In the fruit shop there were many kinds of apples: red, green, big, small, etc. When I came back home I read in the Wikipedia that there are more than 7.500 different kinds of apples: Golden Delicious, Granny Smith, Fuji, McIntosh, etc.

If we were interested in a theory of apples, we would like to know what determines that something is an apple and what determines that something is a Golden Delicious apple and not a Granny Smith or Fuji apple; i.e. what determines that something is an apple at all and what determines that something is the kind of apple it is. The latter may be interesting because we want to sell apples in the American market and we know that the best-sold apples there are, say, Granny Smiths. The former may be interesting, for instance, in a situation in which we arrive to an island and find a tree whose fruit is delicious and we want to sell this fruit in a country where only apples can be sold. Of course, both questions are interesting in their own right.

We are not interested in apples but in phenomenal consciousness. We undergo many different kinds of phenomenally conscious experiences: just consider the experience you have while looking at the ocean, smelling just brewed coffee, when having a headache or having an orgasm. As a theory of apples must explain what determines that an apple is the kind of apple it is and an apple at all, a theory of phenomenal consciousness must explain what determines that an experience is the kind of phenomenally conscious mental state it is and a phenomenally conscious mental state at all.

The previous chapter addressed the first question. I have maintained that differences in the phenomenal character are differences in the content of the mental state.This chapter deals with the second question, what determines that a state is a phenomenally conscious state at all? What is the condition that a state has to satisfy for being a phenomenally conscious mental state?

The phenomenal character of an experience E of a subject S, is the way it is like for S to undergo E. It is in virtue of its phenomenal character that E is the experience it is and a phenomenally conscious experience at all. When I look at my recently bought apple, there is a *redness way it is like for me* to look at the apple. The *redness way it is like for me* to look at the apple is the phenomenal character of the experience, namely PC_{RED}.

PC_{RED} has two different components, the *redness* component and the *for-me* component. Following Kriegel, I have called them qualitative character and subjective character. The qualitative character is what makes the phenomenally conscious experience the kind of experience it is; it distinguishes, for instance, the experience I have when I look at the red apple from the experience I have when I look at a golf-course. In this chapter I will maintain that the subjective character, the *for-me* component, is what makes an experience a phenomenally conscious experience at all. I will argue that phenomenally conscious

experiences have a common component and I will offer an account of it compatible with materialism. In particular, I will argue that a state that has the representational properties discussed in the former chapter, when properly characterized, is a phenomenally conscious mental state.

In section 5.1, I will try to clarify the notion of subjective character to get clear about the phenomenon that we try to explain: all my phenomenally conscious experiences have something in common, a common first-person perspective in which a certain quality is presented to me. I will offer two different, but interrelated, arguments in favor of the subjective character of experience. The first one is based on phenomenological observation. The second one, for those skeptical about phenomenological observation, is based on the analysis of the content of experience from the previous chapter.

My purpose in the rest of the chapter will be to look for a characterization of this property.

In the second section, 5.2, I will argue against theories that try to explain the subjective character of the experience as some form or other of cognitive access. I will discuss two arguments that suggest that a mental state can be phenomenally conscious without thereby being accessible to cognitive processes. The first one, the *mess argument*, is due to Ned Block; the second one, the *dream argument*, is an original one.

The third section 5.3, presents and rejects theories of consciousness that explain the subjective character of the experience as a further representational relation. According to these theories, phenomenally conscious mental states are mental states that are adequately represented.

These representational theories of the subjective character can be divided into two groups depending on whether the mental state is represented by a numerically distinct mental state (higher-order) or not (same-order). Subsection 5.3.1 introduces higher-order theories and some objections they face that lead me to discard them; subsection 5.3.2 introduces same-order theories, particularly Kriegel's proposal, and my reasons for rejecting them as a plausible account of the subjective character.

I will finally present my own proposal in 5.4, this proposal intends to satisfactorily account for the subjective character of the experience without facing the problems of other theories. I will call this theory Self-Involving Representationalism (SIR).

5.1 WHAT IS THE SUBJECTIVE CHARACTER OF THE EXPERIENCE?

Phenomenally conscious experiences are individuated by their phenomenal character: the way it is like for the subject to undergo the experience.

Qualitative properties account for the differences in the phenomenal character of different experiences. The phenomenal character of the experience I have when I look at a red apple and when I look at a golf course are different. Undergoing these experiences is being in two different mental states that have different qualitative properties. In the previous chapter I have argued that the differences in phenomenal character are due to differences in the content represented by the experience. Two experiences have different phenomenal character because they have different content: qualitative properties are representational properties of a particular kind as we have seen.

A theory of phenomenal consciousness has to account, additionally, for the difference between states that are phenomenally conscious and states that are not. A theory of phenomenal consciousness has to explain in virtue of what a state is a phenomenally conscious mental state at all. The mental state I am in while looking at the red apple and when looking at the golf course differ in qualitative properties, but I will argue that they share a property and that it is in virtue of this property that they are both phenomenally conscious mental states. Both experiences, despite having a different qualitative character, have something in common: a common subjective character. All my phenomenally conscious experiences have something in common, a common first person-perspective in which a certain quality is presented to me. The subjective character of the experience seems to me to be self-evident. It is phenomenologically manifest.[1]

Expressing what is phenomenally manifest in ordinary language is a very complicated matter and, unfortunately, some people may find this kind of motivation obscure and suspicious. I do not want to lose readers at this early stage; for that reason, I will offer two arguments in favor of the subjective character. The first one is based on the phenomenological observation. The second one, is based on the analysis of the content of the experience.

5.1.1 *Subjective Character as Phenomenologically Manifest*

The subjective aspect of the experience is a property all phenomenally conscious experiences have in common. In that sense, it accounts for what makes an experience a conscious experience at all. To a first approximation, the best way to point out to this common element is, I think, by examples.

As I held in the introduction, you can distinguish between experiences as of different shades of red, say RED_{35} and RED_{40}. These two experiences are more similar, phenomenologically speaking, between them than with regard to the one I have when I have an experience as of a RED_2. Furthermore, experiences as of RED_{35}, as of RED_{40}, and as of RED_2 seem to be more similar among them that an experience as of GREEN_3. In general we distinguish experiences as of red from experiences as of green.

The phenomenal character of experiences as of red and experiences as of green are in a sense different. But they are in a sense similar (the similarities and differences here are meant to be phenomenological): they are color experiences. They differ in a sense from visual experiences of forms, like a visual experience as of a square. And again, these experiences have something in common, they are visual experiences, and in a sense the *way* they *feel*, their phenomenal character, is similar.

Similarly, auditory experiences of an A produced by a violin are more similar to those produced by a viola than those produced by an electric guitar. The experience of an A played by a violin, and the experience of an A one octave below by the same violin have something in common and all the experiences of the notes of a violin have something in common. All auditory experiences have phenomenologically something in common. Tactile experiences have something in common, the same for auditory experiences, visual experiences, taste experiences, pains,

1 P is phenomenologically manifest if and only if the fact that P obtains can be decided by reflection on phenomenological observation.

orgasms, etc; and all experiences have something phenomenological in common. They are, so to speak, marked as my experiences. Phenomenally conscious experiences happen *for the experiencing subject* in an immediate way and, as part of this immediacy, they are implicitly marked as *my* experience. This is what I call the subjective character of the experience. All these phenomenally conscious experiences have something in common, their distinct first-personal character. All phenomenally conscious experiences have this quality of for-meness or me-ishness.[2]

The idea of qualities of the experience being presented to the subject that undergoes such an experience is introduced by Tyler Burge (2007) as follows:

> The aspects of consciousness in phenomenally conscious states are present for the individual, whether or not they are attended or represented. They are accessible to -indeed, accessed by- the individual. Although they are not necessarily accessible to whatever rational powers the individual has, *phenomenal consciousness in itself involves phenomenal qualities* [qualitative properties in my terminology] *being conscious for, present for, the individual*. They are presented to the individual consciousness. This presentational relation is fundamental to phenomenal consciousness. I think that this relation can be recognized *a priori*, by reflection on what it is to be phenomenally conscious. *Phenomenal consciousness is consciousness for an individual. Conscious phenomenal qualities are present for, and present to, an individual*. (ibid. p.405, my emphasis)

I am going to call *the phenomenological observation*, the observation that in phenomenally conscious experiences phenomenal qualities are presented to the individual of experience, as Burge maintains, or that they are "marked as my experiences" as I presented it in the previous example. The subjective character is the property all and only my phenomenally conscious experiences have that accounts for the phenomenological observation. The subjective character makes an experience a phenomenally conscious experience at all.

The phenomenological observation suggests that a certain form of self is constitutive of the phenomenal character of the experience; in having an experience, a quality is presented to oneself. The content of the experience is not merely that such-and-such is the case, but that such-and-such is presented to myself. As the reader notes, this characterization of the content departs from the characterization I did on page 116 of the transparency of experience. I presented there the transparency of experience as the thesis that in having an experience as of a red apple I do not attribute any property to the experience but

2 Someone could suggest at this point that the subjective character, as I am presenting it, is simply another kind of qualitative character. It should be noted that 'qualitative property' is a technical term introduced to refer to the properties that distinguish the phenomenal character of phenomenally conscious mental states. The subjective character as presented above is phenomenologically manifest. So, if one wants to use 'qualitative property' to refer to what is phenomenologically manifest then one is making a different use of the term.
As I have noted before, if this were a claim about the name it deserves, I still prefer to keep a different name to mark that whereas experiences that have a greenness character have not a redness character, they have different qualitative character, this mineness or for-meness is common to all phenomenally conscious experiences. All phenomenally conscious experiences share a subjective character.

to the apple. Here, I am partially rejecting this claim; I am claiming that in having an experience we do not attribute any property to the experience but we self-attribute certain properties (I attribute centered features to the object of the experience). When I have a phenomenally conscious experience, I thereby attribute a property to myself, in the previous example the property of being presented with a red apple.

We have the strong intuition that our conscious mental states constitute in some important sense a unity, they are not "a mere heap or collection of different perceptions" as David Hume (1739) famously claimed.

Kant argued that in order to account for what I have called the phenomenological observation, for mental representations to be mine, we need to account for a certain sense of self and a certain sense of self-consciousness.

> For the manifold representations, which are given in an intuition, would not be one and all my representations, if they did not *all belong to one self-consciousness*. As my representations (even if I am not conscious of them as such) they must conform to the condition under which alone they can stand together in one universal self-consciousness, because otherwise they would not all without exception belong to me. (B132,B133, my emphasis)

More recently, some philosophers have pointed in a similar direction. Flanagan (1993) has argued that phenomenal consciousness involves some weak sense of self-consciousness, not only in the sense that there is something it is like for the subject to have the experience but also in experiencing my experiences as mine.[3]

Bermudez (1998) has also discussed non-conceptual forms of self-consciousness that are "logically and ontogenetically more primitive than the higher forms of self-consciousness that are usually the focus of philosophical debate" (ibid., p. 274). From the neurological perspective, the idea that a sense of self is required for the experience has been defended by Damasio (2000, 2010) or Pollen (2008).

The phenomenological tradition, to present further examples, contrasts, at least, two forms of self-consciousness: a reflective and a pre-reflective one. In reflective self-consciousness, one has access to oneself in the same sense that one has access to other objects. This object, the self, can be very relevant and probably the most valuable, but it is an object of the experience, as it is the apple or the golf course, and can therefore be distinguished from the experiencing subject, it is a mere *Gegenstand*. On the other hand, in pre-reflective self-consciousness, one is aware of oneself *as the subject of the experience*. Something closer to the idea of pre-reflective self-consciousness is what seems to be constitutive of phenomenal consciousness. Gallagher and Zahavi (2006) present the idea of pre-reflective self-consciousness and the phenomenological observation as follows:

3 As Flanagan notes, the kind of self required is not the elaborate sense of self we usually have in mind, aware of the past and anticipating the future, but rather some kind of primitive process that constitutes the basis for it. The required form of self-awareness should better not require the conceptual abilities necessary for such a narrative self. We want to ascribe phenomenally conscious states at least to some animals and pre-linguistic human babies lacking those abilities. I will further discuss the required sense of self in 5.4.

> There is something it is like to taste chocolate, and this is different from what it is like to remember what it is like to taste chocolate, or to smell vanilla, to run, to stand still, to feel envious, nervous, depressed or happy, or to entertain an abstract belief. Yet, at the same time, as I live through these differences, there is something experiential that is, in some sense, the same, namely, their distinct first-personal character. All the experiences are characterized by a quality of mineness or for-me-ness, the fact that it is I who am having these experiences. All the experiences are given (at least tacitly) as my experiences, as experiences I am undergoing or living through. All of this suggests that first-person experience presents me with an immediate and non-observational access to myself, and that consequently (phenomenal) consciousness consequently entails a (minimal) form of self-consciousness. To put it differently, unless a mental process is pre-reflectively self-conscious there will be nothing it is like to undergo the process, and it therefore cannot be a phenomenally conscious process.
>
> The mineness in question is not a quality like being scarlet, sour or soft. It doesn't refer to a specific experiential content, to a specific what; nor does it refer to the diachronic or synchronic sum of such content, or to some other relation that might obtain between the contents in question. Rather, it refers to the distinct givenness or the how it feels of experience. It refers to the first-personal presence or character of experience. It refers to the fact that the experiences I am living through are given differently (but not necessarily better) to me than to anybody else. It could consequently be claimed that anybody who denies the for-me-ness of experience simply fails to recognize an essential constitutive aspect of experience. Such a denial would be tantamount to a denial of the first-person perspective. It would entail the view that my own mind is either not given to me at all — I would be mind- or self-blind — or is presented to me in exactly the same way as the minds of others.[4]

Some form of self seems to be constitutive of phenomenal consciousness if the phenomenological observation is right. In the phenomenological tradition, this idea has been presented by claiming that in conscious experiences the self is represented by the phenomenally conscious experience not *qua* object (reflective self-awareness) but *qua* subject of the experience, *qua* experiencing thing: consciousness of oneself as subject.

Kant also made a distinction between two kinds of self-consciousness: consciousness of oneself and one's psychological states in inner sense (empirical self-consciousness) and consciousness of oneself and one's states via performing acts of apperception (transcendental apperception 'TA')[5]

4 Not everyone agrees with this part as we will see in 5.3.1. The proponent of the priority of the mindreading abilities over metacognition seems to deny the claim that the kind of access we have to our own mind differs from the access we have to the mind of others.

5 Kant used the term 'TA' to refer to the faculty of synthesis and to refer to what he also referred to as the 'I think', namely, one's consciousness of oneself as subject. The latter is closer to the one I refer to here though I am not committing myself to such a demanding

Despite the fact that 'pre-reflective self consciousness' in the phenomenological tradition and 'TA'[6] in Kant are completely different notions, they both try to account, among other things, for the consciousness of oneself as subject in phenomenally conscious experiences. And this is precisely what a theory of subjective character has to explain: no matter what the primary object of the experience is (say certain features of the apple in my experience as of red apple), the experience is also directed to myself. However, it is not directed to myself as an object but as an experiencing subject. In other words, the content of my experience is not merely that such and such is the case, but that such and such is presented to myself. In phenomenally conscious experiences I do not merely attribute certain properties to the object causing the experience, I attribute to myself being presented with a thing with certain features.

In order to self-attribute certain property a form of self is required. The presence of this form of self in the content of the experience helps explaining the phenomenological observation. I am going to call *for-meness* to the property in virtue of which a mental state satisfies the phenomenological observation, a property that all, and just, phenomenally conscious experiences have: the property in virtue of which they are phenomenally conscious experiences at all.

5.1.2 *Subjective Character as a Common Content*

Some readers may find the phenomenological observation unclear or suspicious. It is hard to express in ordinary language what is phenomenologically manifest. For that purpose I chose to present it by ostension, by pointing to what all experiences seem to have in common. Although I think that a theory of consciousness that does not account for the phenomenological observation is incomplete (precisely because what a theory of consciousness is supposed to clarify is what is phenomenologically manifest), in this subsection I am going to follow an alternative route to support my proposal in 5.4.6. I am going to start from the qualitative character of the experience in the search of a common factor that all phenomenally conscious experience have. I will conclude that accounting for such a common factor seems to require also a form of self.

Undergoing a phenomenally conscious experience is being in a phenomenally conscious mental state. There is a property phenomenally conscious mental states have and other kind of states lack. I am going to call for-meness* this property. Nothing prevents for-meness* from being a disjunctive property. One could consider that phenomenally conscious experiences do not have anything phenomenologically manifest in common and deny the phenomenological observation (I think that this is wrong, phenomenally conscious experiences do, at least, seem to have something phenomenologically manifest in common and probably this seeming to have something in common is all that matters for consciousness. In any case, I am going to concede that for the sake of the argument). There are different ways of being phenomenally conscious, associated with different qualitative characters. Different

notion. Arguably most animal and infants are unable to entertain that kind of 'I thought' and intuitively they do undergo phenomenally conscious mental states.
For a detailed presentation of self-consciousness in Kant's work see Brook (2008).

6 It is unclear how a transcendental posit could explain the appearance of conscious mental unity since that appearance is itself an empirical occurrence as Rosenthal (2005, p. 340) has noted.

ways of being a phenomenally conscious mental state associated with different phenomenal qualities. If someone had this in mind, he should claim something as the following:

(Disjunctive)

A mental state M is phenomenally conscious if and only if M has one qualitative property or other.

According to the proponent of (Disjunctive), phenomenally conscious experiences do not have anything in common. An experience is phenomenally conscious merely if, and only if, it has one qualitative character or other. The phenomenal character turns out to be identical to qualitative character and for-meness* is a highly disjunctive property; namely, having one qualitative character or other.

We can distinguish mental states that are phenomenally conscious from those that are not. The former have a property that the latter lack, namely for-meness*. There is certain condition a mental state has to satisfy in order to count as a phenomenally conscious mental state. The proponent of (Disjunctive), by claiming that a mental state is phenomenally conscious if and only if it has one qualitative property or other, is not saying anything illuminating at all unless she can provide a list with all the possible qualitative properties. I do not think that providing such a list is a very promising project.

A fairer reading of (Disjunctive) claims that qualitative properties have something in common.

(Disjunctive-fair)

A mental state M is phenomenally conscious if and only if M has any qualitative property.

A mental state M has a qualitative property (any) if and only if P.

Where P is the common condition all qualitative properties satisfy. There is a certain condition a mental state has to satisfy in order to fall under the extension of the term 'qualitative property.' In that case, a mental state is a phenomenally conscious experience if and only if it has P. P is for-meness*.

Differences in qualitative character were explained in the previous chapter as differences in the representational content. In particular, I argued that the content of the experience is de se. Before considering this alternative (that for-meness* is the property of having de se content) let me think over a more straightforward theory that would maintain that what all qualitative states have in common is that they have representational content:

(Representationalism)

A mental state M is phenomenally conscious if and only if M has representational content.

(Representationalism) is an interesting proposal supported by the transparency of experience: when I undergo a phenomenally conscious experience as of an apple, I attribute certain features to the apple. However, we ascribe content to some mental states that are not phenomenally conscious. If there are states with content that are not phenomenally conscious, then having representational content does not suffice for having for-meness*. Let me present some evidence in favor of contentful non-phenomenally conscious mental states.

NON-CONSCIOUS REPRESENTATION It is a common assumption in cognitive science that there are non-conscious representations. We attribute to others unconscious beliefs and desires, contentful states. Some of these beliefs and desires are unconscious. For instance, my unconscious desire to kill my father is about my father; it is a contentful state but it is unconscious.

I prefer to focus on non-conscious states that have a content that is similar to the one phenomenally conscious experiences have. Given that I am mainly considering perceptual experiences in my examples I will offer some empirical support to the claim that we have non-conscious states involved in perception. On the assumption that when we perceive that such-and-such is the case we are in a mental state whose content is that such-and-such is the case, then if there are non-conscious states of perception, then there are non-conscious states with content.

The most widely known claim of non-conscious perception was the one made by market researcher James Vicary in 1957. Vicary claimed to have been able to increase the sales of popcorn and Coke by flashing advertising messages like "Eat Popcorn" and "Drink Coca-Cola" during a film whose duration was below the threshold of conscious perception (3/1000 of a second every 5 seconds), so that the patrons did not notice it. Although the weight of evidence suggests that Vicary's claim was in fact a fabrication, there have been numerous studies since then establishing that stimuli can be non-consciously perceived (Merikle and Daneman (1999)).

A classical empirical evidence to support non-conscious experience is blind-sight. Patients with damage in certain brain areas (area VI of primary visual cortex) report blindness in a portion of their visual area. They claim not to be aware of seeing anything in the 'blind' area or scotoma. It has nevertheless been discovered that those patients are surprisingly good when asked to guess about certain objects presented in the 'blind' field.[7] This seems to be a clear support of perception without consciousness. The patient somehow perceives the object, she is in a contentful mental state, but this mental state is not phenomenally conscious.

Another neurological syndrome in which non-conscious perception happens is prosopagnosia or face agnosia. Patients with this syndrome are unable to recognize familiar faces; although they may be aware that they are looking at a face, they are not able to decide who the face belongs to. Some prosopagnosic patients are, nevertheless, able to make correct force choices of the name of the person they claim not to being able to recognize.

In non pathological patients, several studies have shown that there can be perception of information without phenomenal consciousness. For instance, several studies have shown that the orientation of lines (Baker (1937)), geometrical figures (Williams (1938)), or the meaning of words (Merikle et al. (1995)) can be perceived under conditions that do not lead to visual conscious experiences.

In a more recent study on the perception of emotions expressed in human faces, fMRI has revealed that fearful and happy faces lead to differential activation of the amygdala of the observer's brain even in conditions that make it impossible for the participants to explicitly identify the emotion expressed by the presented faces (Whalen et al. (1998)).

7 See Weiskrantz (1986) for a detailed presentation of the phenomenon.

Evidence of nonconscious perception has been reported in different modalities. In smell, Schnall et al. (2008) showed that smells can be non-consciously perceived. They report that unconsciously smelling a fart spray can lead people to make harsher moral judgments they would not otherwise make. In touch, Pagano and Turvey (1998) report that people can determine the length of a wielded object under conditions that prevent conscious experiences of touch like anesthesia or other neuropathology. Whereas it is controversial where to set the threshold for phenomenal consciousness, there seems to be uncontroversial evidence in favor of non-conscious representation.[8]

If there are non-conscious mental states then (Representationalism) is false. A mental state can have representational content without thereby being phenomenally conscious. One way of solving this kind of problems is by denying that having any content suffices for the mental state to be phenomenally conscious; it has to be the right kind of content. In the previous chapter, I argued that the content of the phenomenally conscious experience is something like:

> The centered feature of being disposed to cause experiences with some phenomenal character in me *in normal circumstances*.

This kind of mental states does not merely involve possible world-content but centered-worlds content instead. The content of phenomenally conscious states is *de se*.

With this tool in hand, we can make a new proposal:

> (Self-centered)
>
> A mental state M is phenomenally conscious if and only if M has *de se* representational content.

It seems clear to me that the content of non-conscious mental states in perception is not *de se*. By having an unconscious experience we do not self-attribute any interesting property and there is no need to ascribe de se content to the non-conscious perceptual states of, say, the blind-sighter. Even if any kind of indexical content were required for characterizing the mental states in non-conscious perception this wouldn't be *de se* content.[9]

If this is right, (Self-centered) is progress with respect to (Representationalism), but does not seem to be enough. First of all, if we have non-conscious beliefs and desires, then some of them could be *de se*. Consider for instance my Freudian desire to kill my own father. This desire is non-conscious and its content is plausibly *de se*.

This possible objection can be easily blocked by appealing to the concrete kind of de se content.

> (Self-centered*)
>
> A mental state M is phenomenally conscious if and only if M has the right kind of *de se* representational content.

where the right kind of content has the form: the centered feature of being disposed to cause experiences with some phenomenal character in me *in normal circumstances*.

8 For a summary on non-conscious perception see Merikle and Daneman (1999)

9 See the discussion on an indexical proposal for the content in 4.3. We have no reason for demanding the content of non-phenomenally conscious experiences to satisfy neither POS-SAMENESS nor INCOMPATIBILITY.

(Self-centered*) addresses the problem of non-conscious de se beliefs: the content of these beliefs is not the right kind of content. However, if *the right kind of content* is characterized as above, it seems to be possible to entertain a belief with the right kind of content that is not phenomenally conscious:

> I believe that the object O is causing an experience with phenomenal character PC_1 in me.

When I have this belief, I am in mental state M. I will have this belief even if there is no object causing the experience, but, what is more important, I think that I can have this belief even if I am not having an experience with phenomenal character PC_1. However, M has *de se* representational content of the relevant kind and, according to (Self-centered*), M should be phenomenally conscious mental state. If this is true, then (Self-centered*) is false.

There is, nevertheless, a relevant difference between the content of the belief and the content of the experience. The latter and not the former is non-conceptual.

The idea of non-conceptual content was introduced by Gareth Evans (1982), who maintained that the information tracked by the perceptual system is not organized in concepts. According to the proponents of non-conceptual content, mental states can represent the world even if the bearer of those mental states does not possess the concepts required to specify their content.

It seems clear to me that the content of experience is non-conceptual. Creatures lacking linguistic abilities would probably lack the appropriate concepts, but we do attribute experiences to those creatures like infants and some animals. Furthermore, the content of perception is more fine-grained than the content of propositional attitudes. Evans rhetorically asked: "Do we really understand the proposal that we have as many color concepts as there are shades of color that we can sensibly discriminate?"(Evans, 1982, p. 229) For instance, human beings can discriminate more than 150 different wavelengths, corresponding to different colors, just in between 430 and 650 nanometers. However, if they are asked to reidentify single colors with a high degree of accuracy, they can do so for less than 15 (Halsey and Chapanis, 1951). It is hard to believe that we have the conceptual capacities to make these discriminations. This is even more evident when we think about our olfactory abilities and the conceptual abilities that we have with respect to odors.[10]

I do not even start to find my proposal in the previous chapter plausible if its content is taken to be conceptual; I do not think that having the concepts required for the specification of the content (concepts like PHENOMENAL CHARACTER or CAUSATION) is a necessary condition for having a phenomenally conscious experience. If we grant that the content of perception is non-conceptual, we have a condition that distinguishes the content of phenomenally conscious mental states from the content of the mental states I am in while having a belief or a desire, even if we use the same linguistic expressions for characterizing such content. The content of the experience, contrary to the content of propositional attitudes, is entirely non-linguistic and non-conceptual. This leads to the following proposal:

10 For further arguments in favor of non-conceptual content see Crane (1992); Dretske (1981); Peacocke (1986).

(Non-conceptual Self-centered)

A mental state M is phenomenally conscious if and only if M has *the right kind of non-conceptual de se* representational content.

A mental state is phenomenally conscious if, and only if, it represents the centered feature of being disposed to cause experiences with some phenomenal character in me *in normal circumstances*.[11] Ascribing the object of experience with centered features is equivalent to the self-attribution of properties. The fact that the content of the experience is self-attributive accounts for the subjective character of the experience. For-meness* is the property of representing a certain kind of *de se* content (dispositions to cause the experience in me).

The analysis of the content of phenomenally conscious mental states leads to a conclusion that perfectly matches the phenomenological observation about the relation between the self and qualitative properties. As I noted, this relation can be picked up by maintaining that in phenomenally conscious experiences I do not merely attribute certain properties to the object causing the experience, I attribute to myself being presented with an object with these properties. The content of my experience is not merely that such and such is the case, but that such and such is presented to myself. This is equivalent to the *de se* claim that the object is disposed to cause in me an experience with certain phenomenal character. The content of an experience with phenomenal character PC_1 is the centered feature of being disposed to cause experiences with phenomenal character PC_1 in me in normal circumstances.

In phenomenally conscious experiences I do not merely attribute certain properties to the object causing the experience (as Harman and Tye claim), I actually attribute to myself being presented with a thing with certain features. It seems natural to claim that the conclusion we derived from the phenomenological observation and from the content of experience allows us to maintain that for-meness is for-meness*. However, there is nothing surprising here and this is precisely what we would have expected. Claims about the content of experience are derived from phenomenological observation, from reflection on the phenomenal character of our phenomenally conscious experiences; what would have been surprising is that we would have obtained a different conclusion.

For-meness is the property of representing a certain kind of *de se* content (dispositions to cause the experience in me). The particular content (the particular disposition represented) accounts for the differences in phenomenal character among two experiences.

My proposal is a compresentist inseparatist one according to the taxonomy in 1.3.3. It is compresentist because the phenomenal character is identical with a combination of qualitative character and subjective character and it is inseparatist because the subjective character is a constitutive part of the qualitative character.

So far, I have tried to qualify the explanandum, the task of the following sections will be to provide further clarification of for-meness: the property in virtue of which a phenomenally conscious experience is a phenomenally conscious experience at all. I will try to clarify how a state can have such a content.

11 This proposal seems to be circular, but there is nothing viciously circular in it, as we have seen on page 167.

The subjective character of the experience points toward an intimate relation between the subject and the object of the experience. According to some philosophers, this relation supports the claim that conscious experiences necessarily entail some form of awareness. It seems to be a trivial observation that mental states we are completely unaware of are unconscious mental states. Ned Block presents this suggestion as follows:

> We may suppose that it is platitudinous that when one has a phenomenally conscious experience, one is in some way aware of having it. Let us call the fact stated by this claim – without committing ourselves on what exactly that fact is – the fact that phenomenal consciousness requires Awareness. (This is awareness in a special sense, so in this section I am capitalizing the term.) Sometimes people say Awareness is a matter of having a state whose content is in some sense "presented" to the self or having a state that is "for me" or that comes with a sense of ownership or that has "me-ishness" Block (2007a, p. 484)

Kriegel (2009) points in the same direction:

> [T]o say that my experience has subjective character is to point to a certain awareness I have of my experience. Conscious experiences are not states we may host, as it were, unaware. (ibid. p. 8)

Relying on this fact, some philosophers have investigated this Awareness in the search of a characterization of for-meness: the property all, and only, phenomenally conscious mental states have. I want to maintain that we are Aware of phenomenally conscious mental states in virtue of their having for-meness. But before presenting my own proposal I want to present some objections to other competing theories.

In the next section I will discuss theories that maintain that for-meness is some form or other of cognitive access or accessibility. I will discuss two arguments that suggest that a mental state can be phenomenally conscious without thereby being accessed by any cognitive process. In 5.3 I will discuss and reject alternative theories that maintain that for-meness is a form of representational content. The last section (5.4) presents my own proposal that intends to satisfactorily account for the subjective character of the experience without facing the problems of other theories. I will call this theory Self-Involving Representationalism (SIR).

5.2 SUBJECTIVE CHARACTER AS COGNITIVE ACCESS

Some theories hold that the form of Awareness, to make use of Block's terminology, that is characteristic of phenomenal consciousness is a form of cognitive access. These theories, some way or other, hold the following principle:

> (Cognitive)
>
> Subjective character is a form of cognitive access.

My aim in this section is to show that theories of consciousness that rely on something closer to the (Cognitive) principle are wrong. I will

first present some well-known theories that can be said to rely on, or are committed to, this principle: Tye (1997, 2002)'s PANIC theory, Baars (1988)'s Global Workspace, and Higher-Order theories, particularly the version offered by David Rosenthal (1997, 2005)). I will then present two arguments against these views.

5.2.1 *Cognitive Theories of Awareness*

In this subsection I am going to present theories of consciousness that some way or other endorse (Cognitive). They hold on the idea that phenomenal consciousness is constituted by some form or other of cognitive access.

Tye's PANIC Theory

Michael Tye (1997, 2002) has presented the most developed version of representationalism: PANIC. According to PANIC, the phenomenal character of the experience is given by its intentional content, where 'is given' is to be understood not causally but constitutively; i.e. phenomenal character is constituted by intentional content of a certain kind. Concretely, he characterizes this intentional content as PANIC: Poised, in the sense that it is available to first-order belief-forming and behavior-guiding systems; Abstract, meaning that the intentional content is not individuated by the particular things represented; and Non-conceptual in the sense that it is not structured into concepts.

Granting the possibility of non-conscious abstract and non-conceptual intentional content (a very plausible assumption as we have seen above), Poised is presumably the part of the theory responsible for the distinction between phenomenally conscious states and other kind of states and therefore the part responsible for accounting for the subjective character of the experience. The difference between conscious and non-conscious mental states is a difference in functional role: the former but not the latter is available to first-order belief-forming and behavior-guiding systems. PANIC maintains that the content of the mental state should not be accessed but accesible. According to PANIC, the subjective character is explained by the mere availability to a certain cognitive system. Poised is defined as availability to first-order belief-forming and behavior-guiding systems.

Some philosophers (Burge (1997); Kriegel (2009)) have objected that Poised cannot be a satisfactory explanation of the subjective character. PANIC makes use of a dispositional notion, Poised, to explain something categorical. I think that it would be fairer to read Tye as maintaining that *for-meness* is to be identified with the categorical basis of the accessibility. But, in this case, we have to be told what the categorical basis of the availability is to judge the plausibility of the explanation, and in any case, identify for-meness not with Poised but with the categorical basis of Poised.

Tye's theory is unsatisfactory to say the least. It is obscure how the availability to a certain cognitive system is supposed to give rise to the common phenomenology, the *for-meness* we are trying to explain. Poised is not a good candidate for being identified with *for-meness*, because Poised does not even begin to explain in which sense the conscious

mental states are *for the subject*.[12] Which is the relation between the self and the primary content (the ANIC content)? Is the content de se in virtue of being available to these cognitive systems?

Another theory in the vicinity can help us to fix these problems and provide a suitable answer to these questions. A mental state is phenomenally conscious if and only if it has an abstract non conceptual content (ANIC) and this content is encoded in the global workspace (GWS).[13]

Global Work Space Theory

The Global Work Space (GWS) is a theory developed by Bernard Baars (1988) that distinguishes conscious from unconscious mental states.

GWS is a kind of memory system that can store information from numerous input systems and which is accessible from a large number of cortical and sub-cortical systems. GWS serves as a global broadcasting memory. A memory system to which multiple modules have access.

GWS theory makes sense of the 'theater metaphor'. According to this metaphor, a 'spotlight of selective attention' shines a bright spot on stage; the bright spot reveals the contents of consciousness, actors moving in and out, making speeches or interacting with each other. The audience (modules with access to the GWS) is not selected by the spotlight and remains unconscious, in the darkness, watching the stage. The actors (contents of mental states) can be *seen* by all the public (the content of the GWS is broadcasted to all the sub-systems that have access to the GWS). Also in the dark, behind the scenes, is the director (executive processes) that shapes the visible activities in the bright spot, but is herself invisible. This director is taken to be an equivalent to the required subjective self (pre-reflexive).[14]

Some of the subsystems that access the GWS are inputs, they produce representations stored in the GWS. Other subsystems are consumers of these representations (the public in the metaphor) in the memory system that the GWS is (Baars calls them 'input' and 'receiving assemblies' respectively). GWS contents are proposed to correspond to what we are conscious of, and are broadcast to a multitude of unconscious cognitive brain processes, the consumer systems.

Globally broadcasted messages can evoke actions in receiving processes throughout the brain. The global workspace may be used to exercise executive control to perform voluntary actions.

Allied processors/assemblies compete for access to the global workspace, striving to disseminate their messages to all other processes in an effort to recruit more cohorts and thereby increase the likelihood of achieving their goals. Baars calls these allied processes 'contexts'. Contexts are "coalitions of neuronal assemblies, which can select, evoke, and shape the content of the global workspace, without themselves becoming conscious." Baars (2009)

A stable coalition that routinely controls access to the global workspace is called *dominant context* and, according to Baars, is taken to be equivalent to the subjective self of common sense psychology (the director in

12 I am not demanding an a priori explanation à la Chalmers and Jackson (2001), I am merely claiming that as a theory of subjective character it provides insufficient reasons for an a posteriori identification between poised and for-meness.

13 Note that being encoded in the GWS is an ocurrent property as for-meness is.

14 Baars argues that GWS is distinct from the concept of the Cartesian theater, severely criticized by Dennett (1991), since it is not based on the implicit dualistic assumption of 'someone' viewing the theater, and is not located in a single place in the mind.

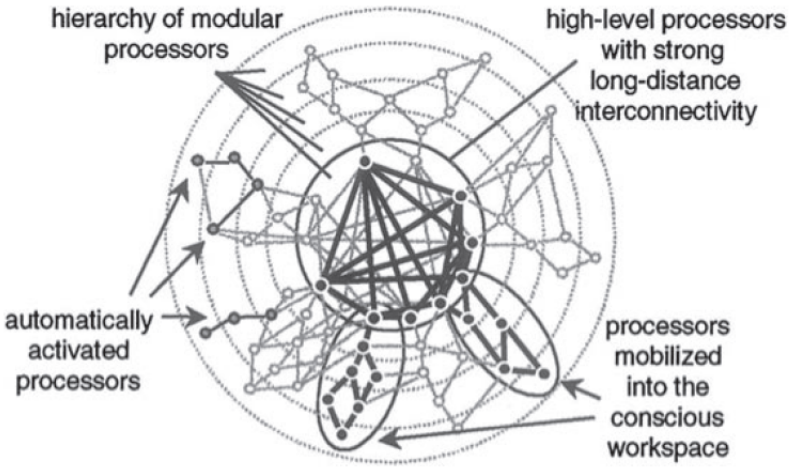

Figure 8: Schematic Diagram of the Global Workspace (Block (2009)).

the metaphor); an 'executive interpreter' in the brain, with certain control of the contents that are made conscious. The 'executive interpreter' is also taken to control voluntary selective attention.

In this case the awareness that would explain *for-meness* would be a cognitive access. The representations are *for the self*, where this self is identified with certain neural assemblies which select and shape the contents of the GWS. The interaction between GWS content, the subjective self and receiving assemblies give rise to conscious experiences, according to the GWS theory. A mental state has for-meness if it is in 'the perceptual space of the self', where the perceptual space is the GWS.

This theory has received important scientific support and popularity in cognitive neuroscience, thanks mainly to the work of Dehaene and colleagues (see Dehaene (2009) for a review), who have provided an impressive collection of evidences to the effect that our ability to report our phenomenal states hinges on such a GWS. According to them, the connection between perception and the workspace lies in long-range neurons in sensory areas in the back of the head which feed forward to the workspace areas in the front of the head. This idea is illustrated in figure 8. The salience of a processor depends on the number and kind of neurons that the process is able to recruit. The concentric circumferences in the picture represent a hierarchy of processors. In the external circumferences are low-level neurons; by that I mean neurons closer in the processing of information to the sensory organ. In the center are high-level ones, closer in the previous respect to cognitive processes. Neurons in a coalition or processor form feed-forward loops increasing the salience of the process by strengthening performance, reducing complexity and often enhancing stability. The possibility of recruiting more processors depends on the salience of the processor. In the figure we can see some automatically activated processors that do not recruit enough cohorts to reach the GWS. On the other hand, we can see processors that are mobilized into the GWS. A central GWS constituted by long-range cortico-cortical connections, assimilates other processes accordingly to their salience. The ones that enter the global workspace are the ones that are conscious according to the GWS theory.

GWS is a very interesting proposal for a characterization of access consciousness, but, as I will try to show, there is empirical evidence

that suggests that it is false as a characterization of phenomenal consciousness. These evidences are reviewed in 5.2.2.

Higher-Order Thought Theory

Higher-Order representational theories commonly claim that a mental state M is phenomenally conscious if, and only if, it is the target of the right kind of higher-order mental state. I will introduce and discuss higher-order theories in more detail in section in 5.3.1. In this section I want to focus on a particular kind of higher-order representational theories. Higher-order thought (HOT) theories maintain:

> (HOT)
>
> A mental state M of me is conscious if, and only if, it is appropriately represented by a higher-order thought.

Probably the best well developed HOT theory is Rosenthal's ((Rosenthal, 1997, 2005)). According to Rosenthal's theory, a mental state M is conscious if, and only if, M is represented in the appropriate way by a higher-order thought (HOT); i.e. for-meness is the property of being represented in the appropriate way by a HOT. The HOT has the content: 'I am in such mental state'. For instance, when I look at the red apple I am in a mental state M with a content like 'red apple.' This mental state M becomes conscious when it is accompanied by a Higher-Order Thought of me to the effect that I am in this mental state; a HOT with a content like 'I see a red apple.'[15] Accordingly, each HOT "characterizes the self to which it assigns its target solely as the bearer of the target state and, by implication, as the individual that thinks the HOT itself" Rosenthal (2005, p.343).

Higher-Order Thought theories endorse (Cognitive), mental states have to be accessed by a higher-order thought to become conscious. In particular, Rosenthal maintains that the cognitive ability underlying reportability is the cognitive ability underlying higher order thoughts. In 'Thinking that one thinks' Rosenthal (2005, chapter 2) writes:

> [G]iven that a creature has suitable communicative ability, it will be able to report being in a particular mental state just in case that state is, intuitively, a conscious mental state. If the state is not a conscious state, it will be unavailable to one as the topic of a sincere report about the current content of one's mind. And if the mental state is conscious one will be aware of it and hence able to report that one is in it. The ability to report being in a particular mental state therefore corresponds to what we intuitively think of as that state's being in our stream of consciousness. (ibid., p.55)

Higher-order thought theory abstracts from any mechanistic interpretation. It could perfectly appeal to something similar to the GWS for that purpose.

15 For Higher-Order Thought theories, phenomenal consciousness requires that the content of the first-order mental state can be re-represented in a thought. Phenomenal consciousness would therefore require that the content of phenomenally conscious experiences were conceptual. I have maintained that this view seems to be implausible. I am going to leave this objection aside, for nothing in the arguments that I am going to present rests on whether the content of the experience is conceptual or non-conceptual.

5.2.2 *Arguments against Cognitive Theories of Awareness*

We have seen some theories that seem to commit themselves to (Cognitive). In this subsection my purpose will be to provide empirical evidence that suggests that phenomenal consciousness does not require cognitive access. If Awareness is required for accounting for the subjective character of the experience then Awareness is not cognitive access.

I will discuss two arguments that suggest that a mental state can be phenomenally conscious without thereby being accessible to cognitive processes. I will first present Block's *mess* argument, and then an original one: *the dream argument.*

The Mess Argument

Ned Block (2007a) has presented an interesting objection to theories that hold on the (Cognitive) principle. He argues that there is empirical evidence that suggests that the content of phenomenal consciousness outstrips our ability to report on such a content. The motivation for this claim is that we are phenomenally conscious of more things than what we can report on. Empirical evidence seems to suggest that our phenomenal consciousness overflows our ability to report it.

Block maintains that we should evaluate theories of consciousness by how well they fit over-all with the body of empirical evidence that we have. The view that phenomenally conscious mental states can occur without cognitive access to the content of these states fits better with these data and (Cognitive) should therefore be rejected. Let me present this evidence.

In a famous experimental paradigm, George Sperling (1960) documented the existence of the iconic memory, a memory buffer with a higher capacity than the working memory.[16] The task of the participants in the Sperling's study were to look at an array of characters (3x4) for a brief period of time, and then recall them immediately afterwards. This technique, called 'free recall', showed that participants were able to, on average, recall 4 to 5 letters of the 12 given. 4 to 5 elements is the capacity of the working memory.[17] However, subjects in the experiment claimed to have 'seen' all of them.

The obvious question that one can raise is the following: did the subjects really *see* all the characters? In other words, is the content of our phenomenally conscious mental states as rich as it seems to be? Or is it just an illusion (the subject can only report 4 items, the capacity of the working memory – the capacity of the GWS)?

Sperling believed that all letters were stored in the viewer's memory for a short period of time, but the memory failed so rapidly that only 4 or 5 could be recalled. Sperling called this bigger buffer iconic memory. In order to test this and whether phenomenally conscious experience persists after the stimulus is turned off, Sperling designed a new experiment. He played tones of different frequencies soon after the blank replaced the array (see Figure 9 for illustration of the paradigm of the experiment). The subjects' task were to report the content of the top row if the frequency of the tone was high, the lower row if the tone was low and the middle row if the tone was intermediate. Subjects were

16 If the global workspace is the categorical basis of cognitive accessibility then it can be identified with the working memory as Block in his argument does.

17 See Block (2007a) for a review of empirical evidence that supports this claim.

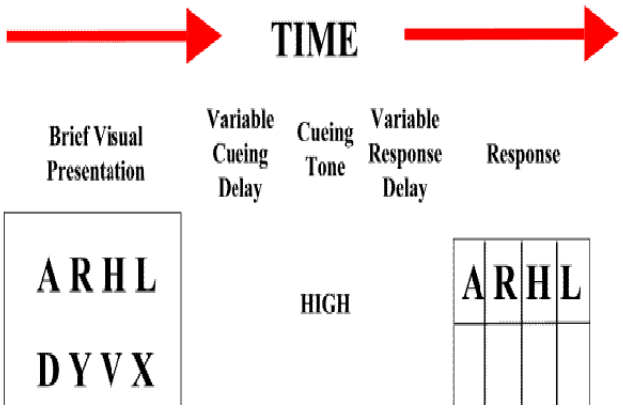

Figure 9: Sperling's Paradigm. Copyright © 2002, Derek J. Smith.

able to report all or almost all of the characters in the indicated row. Given that the working memory can *store* only 4 to 5 elements there must be another buffer with higher capacity where all the letters are stored so that when the cueing tone sounds, the subject can recall the corresponding letters. This is the iconic memory.

There is, however, a further question that remains: is the phenomenally conscious experience persistent or what is persistent is merely the accessible information concerning the stimulus? In other words, are the states encoded in the iconic memory phenomenally conscious? In order to clarify this question, Landman et al. (2003) prepared a test that combines the 'change blindness' paradigm and Sperling's one.

Change blindness is the phenomenon that occurs when a person viewing a visual scene apparently fails to detect large changes in it. In the change blindness paradigm the subject is presented with one image for a short period of time followed by either an identical image or a similar but not identical one with a blank in between, and the cycle starts again. Subjects are often not aware of the changes despite the fact that the cycle is repeated up to 50 times. The effect is widely considered to be an attention-related phenomenon an regarded as inattentional blindness. The best-known study demonstrating inattentional blindness is the Invisible gorilla test Chabris and Simons (1999). Simon and colleagues asked subjects to watch a short video in which two groups of people (wearing black and white t-shirts) pass a basketball around. The subjects are asked to either count the number of passes made by one of the teams or to keep count of bounce passes vs. aerial passes. Someone crosses the scene wearing a gorilla suit and hitting his chest in the middle of the scene. After watching the video the subjects were asked if they saw anything out of the ordinary taking place. In most groups, 50% of the subjects did not report seeing the gorilla. This result suggests that the relationship between what is in one's visual field and perception is based much more significantly on attention than was previously thought. What is under dispute is whether it is a problem of inattentional blindness or inattentional inaccessibility; i.e. whether we fail to perceive the target or we do perceive the target but fail to access the content of the corresponding mental state.

In order to answer this question, Landman et al. (2003) combine the Sperling and the change blindness paradigm. Landman et al. presented the subjects with 8 rectangles arranged in a circle around a point and asked them to keep looking at the central point (see figure 10).

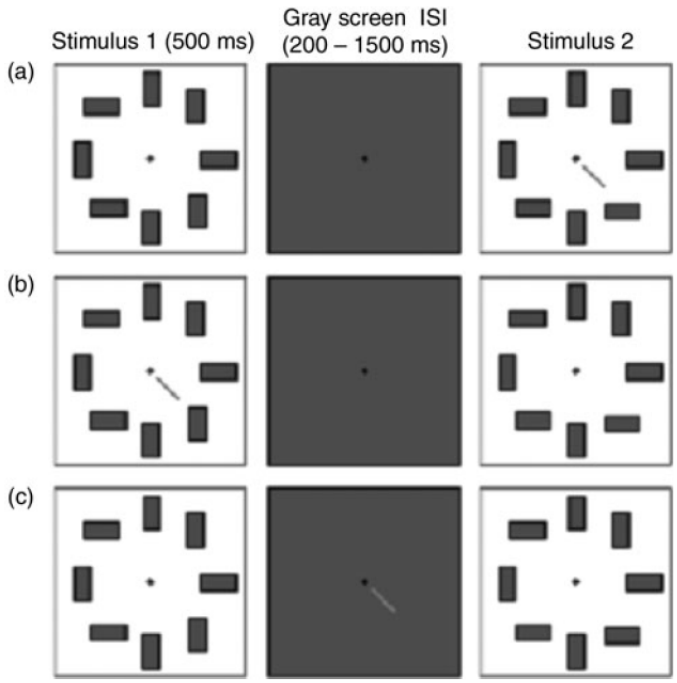

Figure 10: Landman et al.'s Paradigm Landman et al. (2003).

The rectangles could be either horizontally or vertically oriented and were presented to the subjects for 500 ms. The circle of rectangles was replaced by a blank for a variable period (200-1500 ms) and after that by another circle of rectangles in which a line (the cue) pointed to one of the rectangles. The task was to decide whether the pointed rectangle had changed its orientation (Situation (a) in fig. 10). The result was that, after correcting for guessing using statistical procedures, subjects were able to report correctly on just 4 of the rectangles, those in the working memory as we would have expected. Nevertheless, subjects reported having seeing all of them. This is the classic 'change blindness' result and matches up Sperling's result

In a second part of the experiment, the rectangle was cued in the first presentation of the circle (Situation (b) in fig. 10): the rectangle that may change is already cued before disappearing. As expected, the subjects answered almost always right to the question of whether the rectangle had changed orientation.

The interesting result emerged from the third part of the experiment (Situation (c) in fig. 10). In this situation, the line that cues a rectangle appeared during the blank period, after the rectangles had already disappeared. We know that four of the rectangles are stored in the working memory. The subject tries to compare how the rectangles *appear to him* before and after the blank; so, if there were no persistent phenomenology in the iconic memory, then the subject could not compare how the rectangles appear to him before and after the blank, unless the rectangle were stored in the working memory. We would therefore expect a similar result to the one obtained in situation (a). If, on the other hand, there were persistent phenomenology in the

iconic memory then the orientation of the rectangle could be recalled when the cue were presented and we would expect a closer result to the one obtained in (b). The result was that subjects were almost always able to report correctly. Any one of the rectangles was phenomenally accessible when properly cued. This result seems to support the idea that what is both phenomenal and accessible is that there is a circle of rectangles, whereas what is phenomenal, but in a sense not accessible, is the specific orientation of each of the rectangles. There is a sense in which they are accessible, that is, they can be accessed if properly cued and subjects report having seen all of them.

From these results, Block's suggestion is that "the capacity of phenomenology, or at least the visual phenomenal memory system, is greater than that of the working memory buffer that governs reporting" and he uses this fact to argue for the conclusion that "...the machinery of phenomenology is at least somewhat different from the machinery of cognitive accessibility" (Block, 2007a, p. 489).

One could possible reply that there is a generic phenomenology to the effect that there is a circle of rectangles and a specific phenomenology with regard to just some of them and that, when properly cued, the selected rectangle is part of the specific phenomenology. In order to make sense of this, a change in the content of the specific phenomenology should take place (one of the rectangles is replaced by the cued one). One would have to postulate a shift from generic to specific phenomenology when a rectangle is cued. But as Block notes (ibid. p. 532) no subject reports such phenomenological shift. One should expect some change in the phenomenology and there seems to be none.[18]

If we deny that cognitive access is a constitutive part of the phenomenology we can explain the psychological data: subjects report seeing all the rectangles in the Landman experiment – in fact they can report on all of them when properly cued. Cognitive access depends on the working memory, which has a capacity of four items. On the other hand, the iconic memory has a larger capacity, if the iconic memory has phenomenology then subjects can compare how the rectangles appeared to them before the blank and after the blank and we can explain the results of the experiment.

If Block is right, and I think he is, theories that rely on (Cognitive) fail to explain the subjective character of the experience. The Global Workspace theory is a plausible and well supported candidate for explaining the cognitive access we have to our conscious mental states, but not a good candidate for accounting for phenomenal consciousness. The relation between the cognitive processes and the first-order representation is not a constitutive part of phenomenal consciousness.

The Dream Argument

I want to present a second and independent argument against theories that hold on the (Cognitive) principle. More precisely, on theories like HOT that explicitly endorse that phenomenal consciousness constitutively depends on the cognitive access that underlies reportability. We have seen that it is platitudinous that phenomenal consciousness entails some form of *Awareness* as Block (2007a) calls it. Contrary to higher-order theories, first-order theories maintain that Awareness does not depend on the cognitive accessibility that underlies reporting.

18 Further experiments should be performed to test the empirical plausibility of this reply.

As I have tried to show in the introduction, HOT theories maintain that:

A Consciousness requires Awareness;

B Awareness depends on the cognitive accessibility that underlies reporting.

HOT theories are committed to the claim that phenomenal consciousness depends on the cognitive accessibility that underlies reporting. Here is an argument that aims to show that this claim is false.

(DREAM)

(1) The cognitive accessibility that underlies reporting in the case of visual experiences depends on the left dorsolateral prefrontal cortex (dlPFC).

Support for this premise comes from the Lau and Passingham (2006)'s experiment that I will present in short. The conclusion of the experiment is that the neural correlate of the difference between subjects reporting seeing the target stimuli and not seeing it lies in the left dorsolateral prefrontal cortex.

(2) We have conscious visual experiences during the REM phase of sleep.

I will provide empirical evidence to show that:

(3) dlPFC is deactivated during the REM phase of sleep.

(4) dlPFC is not necessary for conscious visual experiences. (From 2 and 3)

∴ Phenomenal consciousness does not depend on the cognitive accessibility that underlies reporting. (From 1 and 4)

Let me discuss the premises of the argument.

THE NEURAL CORRELATE OF COGNITIVE ACCESSIBILITY FOR VISUAL EXPERIENCES: DORSOLATERAL PREFRONTAL CORTEX The evidence for the neural correlate of the cognitive accessibility in the case of visual experiences is based on an experiment performed by Lau and Passingham (2006).

The experiment is based on a visual discrimination task with metacontrast masking.[19] Subjects are presented with two possible stimuli, either a square or a diamond on a black background. After a short variable period of time, SOA,[20] a mask is presented. The mask overlaps with part of the contour of both possible stimuli but it does not overlap with any of them spatially (See Figure 11).

19 In metacontrast masking a second stimulus is presented that interferes with processing and consolidation of the target stimulus in conditions where there is no contour overlap between the target stimuli.

20 The time between the presentation of the stimuli and the mask is called Stimulus Onset Asynchony, SOA.

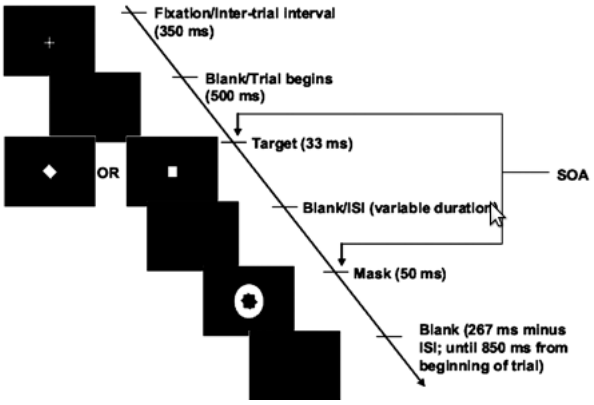

Figure 11: Lau & Passingham's Experimental Set up. Lau and Passingham (2006)

Subjects in the experiment are asked two questions after the presentation of the target and the mask:

1. Decide whether a diamond or a square was presented.
2. Indicate whether they actually saw the target or were simply guessing at their answer.

The first question is intended to measure the objective performance capacity of the subjects. The second question is intended to measure the perceptual certainty of the subjects, how confident they are on having seen the object. This subjective report, according to the author, and to HOT theories, is an indication of phenomenal consciousness.

Figure 12 shows the result as a function of the SOA, the time between the presentation of the target stimulus and the mask. The presence of the mask has nearly no influence on the performance capacity when presented before or close to the stimulus. As the SOA increases, the mask interferes with the target stimulus until it has no effect at all when it is presented much later. The result is a u-shape, where two points with the same performance capacity can be identified.

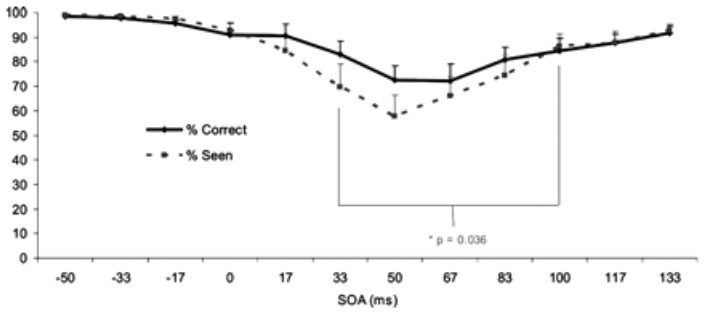

Figure 12: Performance (% correct) vs. Perceptual Certainty (% seen)Lau and Passingham (2006).

The interesting finding for the purposes of my argument is that for some of these pairs of points the perceptual certainty is radically different. Whereas in one (short SOA) subjects report being guessing, in the other (long SOA) subjects are fairly confident of having seen the stimulus. For HOT theories, the subject is phenomenally conscious only in the second case.

Lau and Passingham performed an fMRI study on the subjects of the experiment. Their study revealed that the long SOA condition was associated with a significant increase in activity in the left mid-dorsolateral prefrontal cortex (mid-dlPFC, Brodmann's area 46).

My opponent maintains that Awareness depends on the cognitive accessibility that underlies reporting. In the Lau and Passingham experiment, the subjects report having seen the stimulus in the long SOA condition but not in the short one. Since HOTs are associated with reporting abilities, Lau and Passingham have found the neural residence of HOTs, at least for visual higher-order thoughts ('I see a square').[21] Rosenthal explicitly accepts the evidence from this experiment as showing that the neural correlate of HOTs is in the dlPFC:

> There is, however, some evidence that states are conscious when, and only when, a distinct neural state occurs in mid-dorsolateral prefrontal cortex (area 46) (Lau & Passingham, 2006), and it is reasonable to explore identifying these neural ocurrences with the posited HOTs. Rosenthal (2008, p. 235).

Those who deny (Cognitive) maintain that the curve corresponding to phenomenology could be somewhere in between the two curves in figure 12 (% correct and % seeing) and are not impressed by the fMRI data because they would have predicted exactly this result: the judgment of having seeing, which corresponds to a HOT, is reflected in the prefrontal cortex.

So, does the Lau and Passingham experiment bring some light to the debate between higher-order and first-order theories? I think it does but precisely in the opposite direction from which the authors intended. If HOTs live (or at least a significant part of their neural correlate is) in dlPFC, as the experiment shows, and there were a case of phenomenology without activation of dlPFC, HOT theories would be in trouble. It's time for dreaming.

DREAMS AND DORSOLATERAL PREFRONTAL CORTEX Dreams are the conscious experiences we have during sleep. Revonsuo (2000) defines dreams as '...a subjective experience during sleep, consisting of complex and organized images that show temporal progression'. Dreams are conscious experiences, experiences that are similar in many respects to the ones that we have during wakefulness. Our dreams are highly visual, with rich colors, shapes and movements, and include sounds, smells, tastes, tactile sensations, and emotions, as well as pain and pleasure (Hobson et al. (2000)).

Dreams can be so similar to our waking experiences that the dreamer may be uncertain whether he is awake or asleep. This platitude has been taken for granted by most philosophers. It has, for instance, led philosophers to wonder whether we can distinguish dreams from reality or even whether one could actually be dreaming constantly. This has been referred to by Plato, Aristotle and most famously in Descartes' skeptical argument on the First Meditation.[22]

21 Lau and Passingham maintain that consciousness should be associated with perceptual certainty. Lau (2008) explicitly endorses this view. He maintains that consciousness depends on bayesian decisions on the presence of the stimuli depending on learning processes and the firing pattern of the first-order representations. It is unclear to me why a proposal along these lines should be considered a higher-order proposal. Furthermore, it seems not to be committed to (Cognitive).

22 The common-sense view that dreams are conscious experiences has been explicitly endorsed among others by Kant, Russell, Moore or Freud (Malcolm (1959, p. 4))

I do not intend to argue that dream experiences are exactly like awake experiences.[23] The point that is relevant for the purposes of this argument is that dreams include phenomenally conscious experiences.

Sleep is traditionally divided in two phases: non-rapid eye movement (NREM) sleep and REM sleep.[24] The succession of these two phases is called a sleep cycle. In humans, it lasts for approximately 90–110 minutes; there are 4-5 cycles per night. It has been established that dreams occur during (though probably not exclusively) the REM phase of sleep.

Although there is some controversy as to whether or not there are dreams (or dream-like states) that occur during NREM, there is no much doubt that everybody dreams during REM phase. If subjects are awakened from that stage of sleep and asked whether they have dreamed, they will respond affirmatively at least 80% of the time.

Neurophysiology of REM sleep phase There is a global reduction in metabolic activity and blood flow during NREM sleep compared to resting wakefulness that can reach 40% as shown by positron emission tomography (PET) studies (Braun et al. (1997)). At the cortical level, activation is reduced in the orbitofrontal and anterior cingulate and in particular in the area we are interested in: the dorsolateral prefrontal cortex, Broadmann area 46 (See Braun et al. (1997))

During REM sleep, some areas are even more active than in wakefulness, especially the limbic areas. In the cortex the areas receiving strong inputs from the amygdala like the anterior cingulate and the parietal lobe are also activated (Maquet et al. (1996)).[25] On the other hand, the rest of the parietal cortex, the precuneus and the posterior cingulate are relatively inactive (Braun et al. (1997)).

What is relevant for my argument is that there is no increase in the activity of the dorsolateral prefrontal cortex observed in the comparison of NREM and REM phases. Quite the opposite, both, Braun and Maquet studies, show a decrease in the activity of the dlPFC during REM phase compared to NREM (which, as shown above, presents a reduction in activity with regard to wakefulness). Specifically, Maquet showed a reduction in the area identified by Lau and Passingham for distinguishing subjects that claim having seen the stimulus and subjects that claim being guessing (left dorsolateral prefrontal cortex). All of these regional activations and inactivations are consistent with the differences in mental states between sleep and wakefulness (See footnote 23).

The neural correlate of HOTs lies in the dlPFC; there is an increase in its activity when subjects report having seen the stimuli in comparison with situations in which they report not having seen them and being guessing despite the lack of difference in their performance in both situations. This area is highly deactivated during dreams. If HOTs were constitutive of phenomenal consciousness we should expect a higher

23 According to Tononi (2009, p. 100), dreaming experiences in comparison to waking experiences are characterized by disconnection from the environment, internal generation of a world-analogue, reduction of voluntary control and reflective thought, amnesia and a high emotional involvement.

24 A more fine-grained categorization of the NREM phase can be done based on EEG, EOG, and EMG patterns. For details see Tononi (2009).

25 In the Maquet et al. study, subjects were controlled for dreaming (subjects maintained steady REM sleep during scanning and recalled dreams upon awakening). This control is missing in the Braun et al. study.

level of activity of their neural correlate during REM phase. However, the empirical evidence suggests the opposite.[26]

In what follows I will discuss some possible replies to (Dream) that my opponent can present and my rejoinder to them.

REPLIES

HOTs have a different neural correlate during dreams: One possible way to resist the argument would be to maintain that HOTs have two different neural correlates. (Dream) assumes that visual HOTs have a unique neural correlate. My opponent would claim that during wakefulness dlPFC is the neural correlate for visual HOTs, whereas during sleep HOTs have a different neural correlate.

That kind of dissociation seems, however, implausible. Having another area responsible for HOTs during dreams would require a functional duplication and mutual exclusion. Imagine that we have another area that is the neural correlate of HOTs during sleep,[27] let me refer to this area as 'the sleep neural correlate of HOT '(SNCHOT). When we have a visual experience during wakefulness, the neural correlate of the corresponding HOT is in the dlPFC, and not SNCHOT, which is not differentially activated as the fMRI in the Lau & Passingham experiment shows. During dream experiences, dlPFC is deactivated and the neural correlate of the HOT would be SNCHOT.

REM sleep seems to be exclusive to marsupial and placental mammals (Winson (1993)). It is, therefore, reasonable to assume that the only organisms capable of dreams are those at the top of the pyramid of evolution. The plausibility of SNCHOT depends on the function of dreams during sleep; a function that would require HOTs and would explain the later evolution of an area that fulfills such a function during sleep. If dreams have no function, it seems unreasonable to assume that new changes in brain activity during REM phase appear to give rise to

26 Lau's proposal is not immediately targeted by my argument. If dlPFC is the neural correlate of HOTs, a decrease in the dlPFC activity seems to indicate a decrease in the HOTs entertained and therefore in our phenomenology. On the other hand, for Lau's theory, the role of dlPFC is to work as a Bayesian decision system that tries to make "certainty accurate judgments." The increase in the noise signals (random fluctuations in the neural activity) in the sensory cortex during REM phase in comparison to NREM explains dreams.

> Dreams are more likely to be reported during a stage of sleep that is characterized by rapid eye movement (REM), and brain activity of relatively high frequency and intensity. Let us assume that the overall signal during REM-sleep is higher. If the brain maintains the same criterion for detection over alternations of REM and non-REM sleep, it would be predicted that false positives are a lot more likely during REM-sleep, because of the higher signal intensity.(Lau (2008))

Dreams are for Lau similar to hallucinations, the dlPFC makes the wrong *judgment*. Lau has maintained, in private conversation, that, contrary to HOT, the under-activation of the dlPFC during REM phase is favorable to his theory because in dreams perceptual judgments are wrong. However, in order to properly evaluate Lau's claim we need to be told how the Bayesian decision system is supposed to work and how the decrease of activity in the dlPFC is related to the decision mechanism.
Be that as it may, this reply is not available to HOT theories. In Lau's theory, due to a decreased level of activity, the bayesian decision system is not working properly and makes false positives; it fails to appropriately filter the first-order signals. However, according to HOT theory there is a monotonic relation between the number and complexity of experiences and the number of HOTs and therefore we would expect a higher level of activity in the neural correlate during dreams.

27 A plausible candidate could be the anterior cingulate. As we have seem this area is strongly activated during the REM phase. Furthermore, the anterior cingulate communicates with the relevant sensory and limbic areas.

HOTs in other areas that were not present during wakefulness, and the only area they are present during wakefulness seems to be the dlPFC.

Yet, most of the theories of dreaming yield dreams as epiphenomenal from an evolutionary point of view. (For a review see Revonsuo (2000)) This has been explicitly claimed by Flanagan:

> [Dreams are] a likely candidate for being given epiphenomenalist status from an evolutionary point of view. P-dreaming [phenomenal experiences during sleep] is an interesting side effect of what the brain is doing, the function(s) it is performing during sleep. To put it in slightly different terms: p-dreams, despite being experiences, have no interesting biological function. I mean in the first instance that p-dreaming was probably not selected for, that p-dreaming is neither functional nor dysfunctional in and of itself Flanagan (1995, pp. 9-10; also quoted in Revonsuo (2000) p. 880).

According to the activation-Synthesis theory, Hobson and McCarley (1977), dreams are the result of the forebrain responding to random activity initiated at the brainstem. Dreams are nothing but noise activity.

Other theories either maintain that dreams have a function in memory processing (Crick and Mitchison (1983); Foulkes (1985); Hobson et al. (1994)), in which case there is no function for HOTs and dreams merely reflect the corresponding memory processing (these processes do not require any HOT) or are regarded as some kind of hallucinations that protect sleep without any function for the content of dreams (Solms (1997)).

One exception is Revonsuo (2000).[28] According to him, the function of dreams is 'to simulate threatening events and to rehearse threat perception and threat avoidance'. But this function can also be performed during wakefulness, so the same structures that we use while we are awake could be used during sleep.

As long as one cannot make the case for a function of dreams that would require HOTs, and I seriously doubt that it can be made, we have no reason for defending the possibility of having an additional neural structure, SNCHOT, that differs from dlPFC. There seems to be no reason for a duplication of the HOT machinery. If this is right and dlPFC is the neural correlate of HOTs responsible for visual experiences, then we have good reasons for believing that there are no visual HOTs during dreams.

An alternative objection would deny that we have phenomenally conscious experiences during sleep. This is the next objection I am going to consider.

We do not have conscious experiences during dreams In this case my opponent would reject premise (2). The common sense position maintains that dreams are phenomenally conscious experiences. This position has been endorsed by philosophers, psychologist and neuroscientists, but not without exception.

The common sense position has been famously rejected by Malcolm (1959) who asserts that it leads to conceptual incoherency "... the notion of a dream as an occurrence that is logically independent of the sleeper's waking impression has no clear sense." (op.cit., p. 70). Malcolm maintains that we have no reason to believe the reports given by

28 See also Franklin and Zyphur (2005) for an extension of Revonsuo's proposal.

awakened subjects for there is no way to verify them; these reports could be cases of 'false memory'.[29] It could be that processes during REM phase are all non-conscious and that on awakening there is a HOT targetting the content of memory and thereby making them conscious.

Whereas Malcolm denies that there are dreams, Dennett has defended a skeptical position. Dennett (1976) presents an alternative account in which dreams could be unconscious memory loading processes.[30] According to Dennett, before establishing whether dreams are conscious we need an empirical theory of dreams and it is "an open, and theoretical question whether dreams fall inside or outside the boundary of experience". Dennett goes a step further, claiming that we have some empirical evidence indicating that dreams are not conscious experiences. For instance, he claims that dream activity fails to satisfy well confirmed conditions for phenomenally conscious experience like the activation of the reticular formation (op.cit., p.163).

This position has been challenged by Revonsuo (1995) who provides empirical evidence to the effect that there is in fact activity of the reticular formation and important neuro-physiological similarity between dreaming and wakefulness.

> From the standpoint of the thalamocortical system, the overall functional states present during paradoxical sleep and wakefulness are fundamentally equivalent, although the handling of sensory information and cortical inhibition is different in the two states . . . That is, paradoxical sleep and wakefulness are seen as almost identical intrinsic functional states in which subjective awareness is generated. (Llinas and Pare (1991, p. 522), quoted in Revonsuo (1995))

Unfortunately that would not impress my opponent. According to HOT theory, consciousness necessitates the presence of a HOT; HOTs are absent during dreams, so dreams are unconscious experiences.

Skeptics about dreams base their position on the fact that the access to dreams is retrospective: we recall the dream when we are awakened and we have no reason for trusting these reports, or so the skeptic argues. However, there are cases in which some people are aware of being dreaming. This is the case of lucid dreams. In lucid dreams, the dreamer is able to remember during the dream the circumstances of normal life and to act deliberately upon reflection.

Although lucid dreams have been reported since Aristotle, many have had their doubts about the reality of these episodes. Dennett endorses this skepticism; he considers that the report of lucid dreams is consistent with the subject dreaming that she is aware of being dreaming without any phenomenology involved. But the empirical evidence suggests that Dennett's hypothesis is wrong. The evidence in favor of lucid dreams has been provided by LaBerge and colleagues.

Roffwarg et al. (1962) showed that some of the eye movements of REM sleep correspond to the reported direction of the dreamer's gaze. Based on this evidence, LaBerge et al. (1981) were able to prove the reality of lucid dreams. They trained frequent lucid dreamers and asked them to make distinctive patterns of voluntary eye movements when they realized they were dreaming. These prearranged eye movement

29 Rosenthal, in conversation, points in the same direction.

30 It is not worth discussing the value of the proposal itself, for it is only intended to present a skeptical argument showing that there can be alternative explanations to dreamer's reports when awakened.

signals were recorded by the polygraph records during REM, proving that subjects had indeed been lucid during uninterrupted REM sleep. This result has been replicated by other laboratories. (For a review see LaBerge (1988)).

The experiments on lucid dreams provide evidence that we have conscious experiences during sleep, and give us the opportunity to record reports to that effect. The main reason for skepticism is dissolved: there are conscious dreams.[31]

My opponent can still try to resist the argument by maintaining that we have conscious experiences during lucid dreams but not during ordinary dreams, for only during lucid dreams can the subject report on them (according to her, reporting is inextricably linked to HOTs). One could also claim that the subject having a lucid dream is reporting being dreaming and therefore having phenomenally conscious experiences, but maintain that the subject is not having a phenomenally conscious *visual* experience and Lau and Passingham's experiment merely shows that the dlPFC is the neural correlate of *visual* HOTs.

Let's consider two possibilities: the dlPFC is activated during lucid dreams (the most plausible option) or it is not.

If it is not activated then the subject might be reporting having phenomenally conscious experiences but not a visual one. I fail to see what kind of experiences the subject might be having. In any case, in order to settle this discussion, subjects in the experiment could be asked to move their gaze when they "see something" during lucid dreams. If there were no activation of the dlPFC in this condition, then HOT would be safe. This result would show that the dlPFC is not the unique neural correlate of HOTs, because HOTs are required for reporting and, in this case, the subject would be reporting "seeing something". I have argued that this option is not very plausible.

On the other hand, it might be that dlPFC is activated during lucid dreams. For different reasons, most scientists expect an activation of the dlPFC during lucid dreams.[32] If this is the case, then the only option available for HOT theories is to maintain that we are only phenomenally conscious during lucid dreams. This half-baked reply distinguishing lucid dreams from other dreams in this respect seems to be something of a reach.

Let me sum up the argument in this section. Lau and Passingham's experiment provides good evidence for believing that the neural correlate of the reporting access to our visual conscious experiences depends on the dorsolateral prefrontal cortex which is deactivated during

31 Manolo Martínez has suggested to me an interesting case: sleep talkers. A sleep talker is someone who talks during sleep. I do not know any experiment with sleep talkers so what I will say in what follows is not scientifically supported. There is a coherency between what sleep talkers say while they are sleeping and what they report as having been dreaming about. That suggests that the sleep talker was having a dream during sleep and I think that the dlPFC will be activated. Again the defender of HOT could maintain that the sleep talker is having dreams but not a non-sleep talker; this position is, I think, really unsustainable, for the same reasons that we are going to see.

32 While Tononi (2009) considers this possibility, his motivation is nevertheless different. For him, dreams are conscious experiences normally characterized, among other things, by a reduced voluntary control and reflective thought. Tononi explains this characteristic by the deactivation of dlPFC which is involved in volitional control and self-monitoring. For that reason, Tononi asserts:

> It is plausible, but not proven, that the deactivation of dorsolateral prefrontal cortex that is generally observed during REM sleep may not occur during lucid dreams.

According to Tononi, we should expect an activation of the dlPFC when lucid dreams are reported.

dreams. The evidence seems to suggest that access is not necessary for consciousness, for we lack it during dreams when we are conscious.

I have argued that we have no reason to believe that this function is implemented by other areas during sleep.

The defender of HOT theory can embrace a skeptical position as to whether we have conscious dreams. This position, which runs against common sense, has been refuted by strong empirical evidence (lucid dreams).

The position remaining for HOT theory is not a comfortable one, or so I have tried to argue. If dlPFC is activated during lucid dreams, HOT has to maintain an ontological dichotomy with regard to dreams (some dreams are phenomenologically conscious and others are not). If it is not, HOT theory is seriously jeopardized (unless it is also deactivated when the subject is reporting having a phenomenally conscious experience during the lucid dream).

5.3 SUBJECTIVE CHARACTER AS REPRESENTATION

In the previous section I have presented and rejected theories that appeal to some form of cognitive access for explaining the subjective character. In this section I will consider theories that maintain that a mental state M is phenomenally conscious if, and only if, it is *adequately represented*. Different theories spell out the notion of 'adequately represented' in different ways.

Phenomenally conscious mental states differ in an interesting way from other mental states. At least part of what makes them different is the relation between the individual that holds the phenomenally conscious mental state and the phenomenally conscious mental state. The relation between the subject and the content of the phenomenally conscious mental state differs from the relation that holds between the subject and the content of other mental states. All mental states that I have are my mental states but phenomenally conscious mental states are presented to or conscious for-me. Whereas all mental states are mental states of mine, states that I *host* as I *host* my heart or my kidney, only phenomenally conscious mental states are for-me. Only phenomenally conscious mental states have subjective character.

> If there is something that makes a conscious experience "for me," then by having the experience, I must be somehow aware of having it. For if I am wholly unaware of my experience, there is no sense in which it could be said to be "for me.
>
> The awareness in question is quite special, however. On the one hand, it must be conceded that we are aware of our conscious experiences. For conscious experiences are not sub-personal states which simply happen in us, without our being aware of them. Kriegel (2005, p. 25)

By having a phenomenally conscious experience I *feel* something. A characterization of this requires that the qualities of experience are presented to oneself, that somehow the subject of the experience is *Aware* of them. Following the idea that the subjective character is a form of awareness, some philosophers have proposed that subjective character requires a further representation in which the state that has qualitative properties is represented. I want to show that these theories

fail to explain the subjective character of the experience for different reasons.

In this section, I am going to first motivate theories of consciousness that explain the subjective character of the experience as a further representation relation; phenomenally conscious mental states are mental states that are adequately represented; i.e., for-meness is the property of being adequately represented.[33] Then I will offer some arguments against them.

In the presentation of representational theories of subjective character I will be following Kriegel (2009, ch. 4), where I have found the most clear argument in favor of representationalist approaches to subjective character. I will call this main argument REPRES.[34]

Representational theories of subjective character can be divided into two groups depending on whether the mental state is represented by a numerically distinct mental state (higher-order) or not (same-order).[35] Subsection 5.3.1 introduces higher-order theories and present some objections they face that lead me to reject them. Subsection 5.3.2 introduce same-order theories, particularly Kriegel's proposal, and my reasons for rejecting it as a theory of the subjective character.

The argument in favor of representational theories of subjective character goes as follows:

(REPRES)

The first part of the argument goes from subjective character to awareness.

(1) $\Box[\forall M(Consc(M) \leftrightarrow \exists S(C(S,M) \wedge Subj(M))]$

(2) $\Box[\forall M(Subj(M) \rightarrow \exists S(C(S,M) \wedge Aware(S,M))]$

(3) $\Box[\forall M(\exists S(C(S,M) \wedge Aware(S,M)) \rightarrow Subj(M))]$

---

(4) $\Box[\forall M(Consc(M) \leftrightarrow \exists S(C(S,M) \wedge Aware(S,M))]$ From 1, 2 and 3.

Where S is a subject, M is a mental state, the box indicates metaphysical necessity, C(S,M) indicates that the subject S is in a mental state M, Consc(M) that M is phenomenally conscious, Subj(M) that M has subjective character (is for S) and Aware (S,M) that S is aware of M in *the right way.*

(1) claims that a mental state is conscious if and only if it has subjective character. A *prima facie* problem is that it requires the existence of a subject; but the notion of subject used in the conditional is left unexplained. What are the conditions for having a subject? Is the subject

33 I have already introduced one of these theories: HOT theories. The arguments presented in the previous section target only a particular higher-order theory: HOT theory. In these section I will target higher-order theories in general.

34 Kriegel (2009) calls his argument the Master Argument for the necessity of Self-representationalism. Kriegel's argument presents only necessary conditions for phenomenal consciousness. REPRES is a more detailed formalization and extension of part of this argument. In section 6 of chapter 4 he argues that self-representationalism, the theory he is proposing, provides sufficient conditions for phenomenal consciousness. I will present self-representationalism in 5.3.2.

35 For examples of higher-order theories see Amstrong (1968); Carruthers (2000); Gennaro (1996); Lycan (1996); Rosenthal (1997, 2005); Van Gulick (2004). For examples of same-order theories see Burge (2007); Brentano (1973); Caston (2002); Kriegel (2009)

the same thing as the organism? Can any organism be a subject? I will come back to these questions.

(2) claims that some kind of awareness is a necessary condition for subjective character. The claim that some form of awareness is essential to subjective character seems to be uncontroversial. Different theories offer different characterizations of *being aware in the right* way. The soundness of the argument depends on this characterization and this is what I am going to discuss in the remaining of the section.(3) claims that some kind of awareness is a sufficient condition for subjective character. This claim seems more controversial. The phenomenological observation motivates (2), but not (3), as the quote above makes clear. Whereas (2) seems hard to reject, it is not clear that we have good reasons for accepting (3). If we accept that in having a phenomenally conscious experience the subject is somehow aware of the qualities of the experience we are committed to (2) but not to (3). However, I will accept (3) as an hypothesis for the sake of the argument.

The second stage goes from awareness to representation.

(5) $\Box[\forall X \forall S(Aware(S,X) \leftrightarrow R_1(S,X))]$

(6) $\Box[\forall X \forall S(R_1(S,X) \leftrightarrow \exists M^*(C(S,M^*) \wedge R_2(M^*,X))]$

---

(7) $\Box[\forall M(Consc(M) \leftrightarrow \exists S \exists M^*(C(S,M) \wedge C(S,M^*) \wedge R_2(M^*,M))]$
From 4, 5 and 6.

Where X can be any entity, M^* is a mental state and R_1 and R_2 are representational relations.

(5) claims that being aware of something is a matter of representing it. I take that to be an uncontroversial claim.

(6) explains how a subject can represent something: a subject represents something in virtue of being in a mental state that represents the entity in question. I think that it is important to distinguish the representation relation that holds between the subject and the mental state and the one that holds between mental states, as we will see during the objections to representational theories of subjective character. For that purpose, I have distinguished between two representational relations, R_1 and R_2, to make clear the distinction. R_1 is a representational relation between the subject and the object that is explained as the subject being in a state that is representationally related to the object (R_2).

(7) is the conclusion that a mental state M is conscious if and only if it is represented by a mental state M^*.

The relation between M and M^*distinguishes higher-order theories (HOR) from same-order theories (SOR) of consciousness. HOR theories maintain that M and $M^*$are different states, whereas SOR theories maintain that they are the same.[36]

The remaining of the section is organized in two subsections. In the first one I will present HOR theories and three objections to them. In the second one I present SOR theories, in particular self-representationalism, and offer my reasons for not endorsing this theory as a theory of subjective character.

36 Prima facie the idea of a state representing itself may seem disconcerting. Kriegel's proposal is that a state represents itself indirectly in virtue of one part of it representing the other. I will offer the details in 5.3.2.

5.3.1 *Higher-Order Representational (HOR) Theories*

For HOR theories consciousness should be explained at a certain cognitive level. In the previous section we have seen some evidence that, as I have tried to show, suggests that phenomenal consciousness should not be explained at the cognitive level. I am going to leave these considerations aside in this section; the objections I will present here are independent of the success of the previous arguments.

HOR theories try to explain the subjective character of experience, the difference between conscious and non-conscious experiences. Whereas strong representationalism, as a qualitativist theory, reduces consciousness to a certain sort of intentional content, according to HOR theories first-order content does not suffice for phenomenal consciousness. Phenomenally conscious states are the objects of some kind of higher-order process or representation. There is something higher-order, a meta-state, on the case of phenomenal conscious mental states, which is lacking in the case of other kind of states. HOR theories commonly claim that a conscious mental state is the object of a higher-order representation of some kind. It is in virtue of this higher-order representational content that the mental state is for-me. For-meness is the property of being represented by a higher-order state.

The kind of representation that is required by the theory makes a basic difference among different HOR theories. The main concern is whether higher order states are belief-like or perception-like. The former are called Higher-Order Thought (HOT) theories (Gennaro (1996); Rosenthal (1997, 2005)) the latter Higher-Order Perception (HOP) or 'inner-sense' theories (Amstrong (1968); Carruthers (2000); Lycan (1996)). According to the former theories, when I have a phenomenally conscious experience as of red I am in a mental state with certain content, call this content RED. For this mental state to be phenomenally conscious, there has to be, additionally, a higher-order thought targetting it, whose content is something like 'I am seeing RED.' On the other hand, HOP theories maintain that what is required is a (quasi-) perceptual state directed on the first-order one. Some kind of monitor system that marks some mental states as 'mine.'

A second point of disagreement is whether a given state is conscious in virtue of its disposition to raise a higher-order representation (Carruthers (2000)) or by being actually the target of a higher-order representation (Rosenthal (1997, 2005)); this is the difference between dispositional and actualist HOR theories. According to dispositional HOR theories, the higher-order representation that renders the Awareness of the first-order one doesn't have to be actual, there is no need for the higher-order representation to happen actually, what is needed for a mental state to be conscious is a disposition to be the object of such a higher-order representation. As we saw in the case of Tye's PANIC, for-meness should in any case be identified with the categorical basis of this disposition.

What is relevant for our discussion in this section is that all HOR theories commonly maintain that M and M*are different states. M is phenomenally conscious in virtue of there being another mental state M* that represents it. This is what accounts for the difference between states that are phenomenally conscious and states that are not.

Objections to HOR Theories

I am going to present three objections to HOR theories. I will first argue that it is obscure how they are supposed to account for the subjective character as I have presented it; then the second objection will target a particular theory of consciousness, Carruthers (2000), that is committed to the idea that metacognition depends on mindreading capacities; this objection can be extended under plausible assumptions to other HOR theories. I will finally argue that HOR theories in general are jeopardized by the possibility of a HOR misrepresenting the first-order mental state.

EXPLAINING FOR-MENESS Even if one concedes that some form of HOR theory satisfactorily distinguishes between conscious and non-conscious mental states, it does not explain the subjective character. Let me elaborate:

The existence of *the subject* is a pre-requisite to REPRES. It is unclear what kind of entity this required *subject* is. At the very least, a subject must be something that can hold mental states, because a condition for the conclusion is that both M and M*are *in S* $(C(S, M) \wedge C(S, M^*))$. In such a case, the subject is the raw bearer of mental states.

S is aware of M if and only if S represents M $(R_1(S, M))$. But S representing a mental state is just a matter of S having another mental state, M^*, such that M^* represents M; i.e. $R_2(M^*, M)$. Understood that way, the conclusion of the argument (7) seems to be false. Being represented by another mental state is insufficient for the phenomenal state to become phenomenally conscious. As we have seen on page 179, we have good reasons for maintaining that there are contentful mental states that are not phenomenally conscious; it seems reasonable to assume that these mental states are re-represented, they are represented by other mental states, in further processes like non-conscious beliefs, or more complex motor control, without thereby becoming phenomenally conscious. Consequently, being represented by a higher-order mental state is not a sufficient condition for a mental state to have for-meness. Furthermore (7) doesn't seem to do justice to the phenomenological observation unless R_2 somehow involves the subject of the experience.

Proponents of one or other version of higher-order representationalism acknowledge that and make it clear that the mental state has to be represented *in the right way* in order to become a phenomenally conscious mental state.[37]

I know of two different, but not incompatible, ways in which a higher-order representation could try to explain the subjective character. One corresponds to HOT theories and the other to HOP theories.

According to HOT, a mental state M of mine is conscious if and only if it is accompanied by a higher-order thought to the effect that *I myself* am in M. For HOT theories a mental state with the content RED_{34} becomes phenomenally conscious when it is the target of another mental state with the content 'I see RED_{34}'. A concept of the self is part of the content of the higher-order thought and that way for-meness can be explained. One problem for this proposal independently of the one presented in 5.2.2 is that the kind of content that explains the subjective character is conceptual and as I have maintained on page 181,

37 One could claim that R_2 is a relation between conscious states, but this would lead to an infinite regress, as is well known from the literature (Caston (2002); Kriegel (2009)).

the content of phenomenally conscious experiences should better be non-conceptual.[38]

Proponents of higher-order representational theories can, however, hold on a HOP theory. According to HOP theories, the mental state is represented by some kind of higher-order monitoring system and become thereby conscious. Proponents of HOP owe us an explanation of which is the system and why this monitoring system is relevant in such a way that mental states that are represented by this higher-order monitoring system become conscious. A possible answer at the implementation level is the GWS, where the self is identified with certain 'executive assemblies' that control the access to the global workspace, in that sense representations encoded in the GWS are *for the self*. But we have seen that there are good reasons for resisting the idea that being encoded in the GWS is a necessary condition for phenomenal consciousness: there is empirical evidence that suggests that there are mental states that are phenomenally conscious but are not encoded in the GWS.

An interesting alternative proposal has been presented by Carruthers (2000), this theory is the target of the next objection.

MINDREADING FIRST Carruthers (2000) argues in favor of a dispositional HOR theory. I am going to first present this theory and then present an objection to it. This objection can be extended to all higher-order theories if certain ideas about the relation between metacognition and mindreading are true.

Carruthers maintains that we have first-order perceptual states, these states are representational but not phenomenally conscious states. He explains the contents of mental states by appealing to *consumer semantics*.[39] Consumer semantics maintains that the content of mental states depends on the powers of the system that 'consumes' that state; we have seen some of these consumer theories in the previous chapter.[40]

According to Carruthers, some of these mental states acquire at the same time a higher-order content in virtue of their availability to another consumer system: a theory of mind, the ability of humans to identify their own mental states and ascribe mental states to others. It is in virtue of their availability to the Theory of Mind faculty, as a consumer system, that the perceptual states in question acquire a dual content. Mental states with this dual content are phenomenally conscious mental states. Certain mental states are recognized as mental representations by the Theory of Mind, and this gives them their subjectivity. These representations are dual in content:

> Each phenomenally conscious experience has its distinctive form of subjectivity by virtue of acquiring a higher-order analogue content which precisely mirrors, and represents as subjective, its first-order content.(Carruthers, 2000, p. 243)

Each experience would, at the same time, be a representation of some state of the world (for example, a representation as of red) and a representation of the fact that we are undergoing just such an experience

38 For a further argument against HOT theories based on memory constrains see Metzinger (2003).

39 See for instance Millikan (1984); Peacocke (1995)

40 In particular, Carruthers endorses an inferential role semantics (Block (1986); Peacocke (1995)), according to which the content of a state depends on the kind of inferences which the cognitive system is prepared to make in the presence of the state.

(a representation of *seems red*), through the consumer system that is the Theory of Mind. The concepts produced by the Theory of Mind could make use of first-order representations. This new content, *seems red*, is a by-product of a mind reading faculty, which builds up a distinction between how things are and how they seem (the *is/seems* distinction).

Our evolutionary ancestors would have had first-order representation concepts for many features of the environment (red, green, etc); then the development of a theory of mind would have allowed them to build an *is-seems* distinction. Higher-order recognitional concepts (seems red, seems green, etc.) could have been generated in response to the very same perceptual data that gave rise to the first-order concepts. In the example of an experience as of red, besides there being a first order representation of redness, there is also second-order representation of seeming-redness.

According to Carruthers, the subjective character of the experience is explained as an additional intentional content produced by the Theory of Mind. Carruthers introduces an interesting proposal defending higher-order representational theories of consciousness.

> The explanation of phenomenal consciousness which I am putting forward, then, claims that it is because the content of C is available to two sets of consumers –first order conceptual reasoning systems, as well as a higher-order mind-reading faculty– that those contents actually (categorically) have dual representational status. (ibid., p.246)

A conscious mental state has a double content (*is/seems*) due to these two systems. The second content, provided by the theory of mind, plays the role of explaining what I have called subjective character of consciousness. A given experience is *for-me* if it has the *seeming* dimension. For-meness is the property of being available to a theory of mind. This proposal, while compelling, faces, I think, a serious objection.

In a nutshell, my objection is that consciousness seems to be a prerequisite to a Theory of Mind and not a by-product of it, as Carruthers maintains. I need to know *what-it-is-like for-me* to see red to infer that *what-it-is-like* for the others to see red is something similar and use this information for planning, fooling, etc. Let me present the objection with a bit more detail.

A theory of mind is the ability to attribute mental states like beliefs, desires, etc. to oneself and to others and to understand that others have different beliefs or desires from those that one has. A theory of mind can be decomposed into two abilities: metacognition and mindreading. Metacognition is the ability to attribute to ourselves mental states and mindreading the ability to attribute mental states to others.

Higher-order representational theories commonly hold that metacognition, access to some of our mental states, is a necessary condition for phenomenal consciousness. Beings lacking metacognition lack thereby phenomenal consciousness. Carruthers further claims that the ability of mindreading is also required.[41]

41 Some philosophers consider this to be a reason for rejecting these theories. It is too demanding, for it requires precisely a Theory of Mind and most animals and arguably human babies lack it. I do not consider this last point to be a defeating one. Maybe animals and babies lack phenomenally conscious states after all. Although intuitively they have phenomenally conscious experiences, I can only be sure that I do have conscious mental states and I have no serious doubts that so does the reader. I do not think that a

Although it has been suggested that mindreading and metacognition are two different mechanisms (Nichols and Stich (2003)), it is commonly held that there is a unique mechanism for both abilities and that they are directly connected. There is, however, a huge controversy on whether metacognition is prior to mindreading (where metacognition being prior to mindreading means that the ability of mindreading depends on the mechanisms that evolved for metacognition) or the other way around.

Goldman (2006) suggests that metacognition is prior to mindreading. The attribution of mental states to others depends upon our introspective access to our own mental states together with processes of inference and simulation of various sorts, where a simulation is "the process of re-enacting or attempt to re-enact, other mental episodes". This is what is known as simulation theory of mind. An example by Goldman may help to illustrate the idea:

> Seated in my living room on a wintry day, I might imagine myself instead watching the surf on some sandy beach. What I am trying to do is undergo a visual experience that matches (as closely as possible) a visual experience I would have if I really were on the beach. Vision science tells us that what transpires in visual cortex when undergoing visual imagery can, to a considerable extent, match what goes on during genuine vision (Kosslyn and Thompson, 2000). This is what we call a mental simulation. This is a case of intrapersonal simulation: trying to re-enact an event in one's own mind. In using simulation to read others' minds, however, one would try to re-enact their mental states. That's just how mindreading characteristically takes place, according to simulation theory (ST).Goldman and Shanton (2010)

The opponent to the simulation theory is known as theory-theory. Theory-theory holds that when we mindread, we access and utilize a theory of human behavior represented in our brains. It posits a theory of human behavior commonly known as 'folk psychology.' Just like other folk theories, such as folk physics, it helps us to master our daily lives successfully. On this view, mindreading is essentially an exercise in theoretical reasoning. When we predict behavior, for example, we utilize folk psychology in order to reason from representations of the target's past and present circumstances and behavior (including verbal behavior), to representations of the target's future behavior. For theory-theory, if there is just one mechanism, then metacognition depends on mindreading. Metacognition is merely the result of turning our mindreading capacities upon ourselves. In metacognition we just self-interpret ourselves. This is the view defended by Carruthers.[42]

Carruthers holds on a theory-theory approach to the Theory of Mind (in opposition to a simulation theory). I do not have any defeating argument against theory-theories, what I find implausible is the devel-

theory that maintains that animals and babies are non-conscious is immediately wrong, but surely, when comparing alternative theories, one that doesn't have this consequence is to be preferred.

42 More precisely, in Carruthers (2000), where he presents his theory of phenomenal consciousness, he suggests that mindreading and metacognition are a unique mechanism with two different modes of access, one for perception (mindreading) and one for introspection (metacognition). In Carruthers (2009) he gives up this view, in favor of the one presented here.

opment of a mind-reading faculty without knowing what it is like for the subject to have *any conscious experience*.

My purpose in this objection is to show the implausibility of a theory of consciousness according to which having a phenomenally conscious experience depends on having a theory of mind.[43] This objection may be extended to any higher-order theory that makes mindreading prior to metacognition. In other words, either metacognition is prior to, or an independent mechanism from, mindreading, or higher-order theories face serious problems.[44] The reason is that phenomenal consciousness is a necessary condition for attributing to others mental states that feel some way or other.

My opponent would argue that conscious experiences are not necessary for developing a theory of mind along these lines: creatures can see objects in the environment and the response of other organisms to those objects and their properties. Different properties cause different responses in different creatures. On that basis, organisms (through evolution) can come to theorize that there are internal states inside of other creatures that track particular properties and conditions. Similarly, my opponent would argue, in the case of experiences, when people attribute to others sensory states there is no reason for attributing feeling, we just attribute to them states that track certain properties.

That seems to me to be completely misguided as it is dramatically clear in the case of pains or orgasms. The kind of mental state ascription mentioned above is very different from the kind of attribution we usually do. How can one ascribe others with mental states that *feel* in a certain way if one has never been in a mental state that *feels*? It seems to me that the kind of attribution would be completely different in this case.[45] For illustration, consider Sally who has never had an orgasm in her life. Sally knows that she has never had an orgasm. She can nevertheless ascribe orgasms to other people, as a matter of fact she is really good in that task and she can always recognize when her partners are having an orgasm or just faking given their behavioral response. Surely the kind of mental state Sally attributes to her partners or, for instance, actors when seeing a film, is a phenomenally conscious mental state. My intuition is that clearly, after she has an orgasm for the first time, the kind of experience that she will be attributing to others when having an orgasm is different from the one attributed before she felt an orgasm for that first time. She knows how it feels to have an orgasm and attributes to others a similar sensation when they are having an orgasm.

This example suggests that the kind of mental state attributions that someone that lacks phenomenal consciousness can do, in case she can, are different from the ones that I can do. If this is right, then phenomenal consciousness cannot depend on mindreading capacities, for phenomenal consciousness is prior, at least to certain mindreading capacities. We attribute to others phenomenally conscious mental states and this kind of attribution is not possible unless one has undergone the relevant experience, as the example suggests. So, phenomenal consciousness cannot be a by-product of our mindreading capacities,

43 In fact Carruthers (2009) seems to take feelings as inputs for a mindreading ability.

44 For a deep discussion in favor of the priority of mind-reading to metacognition and replies see Carruthers (2009).

45 This is independent of whether my ascription of mental states to myself or others is due to a simulation theory or purely theoretical.

precisely because our mindreading capacities require phenomenally conscious mental states.

One possible alternative theory, not clearly in the spirit of Carruthers' one, would maintain that our theory of mind evolved in two steps. In a first step a proto-theory of mind attributes states with certain functional role. A mental state is phenomenally conscious in virtue of being available to this proto-theory of mind. In a second step a full-blown theory of mind evolves and allows the attributions of phenomenally conscious states to others.

The problem of this reply is that, according to Carruthers, the functional role attributed by the proto-theory of mind exhausts the phenomenal character and the proto-theory of mind already allows the attribution of mental states with this role. So, there is no evolutionary advantage in attributing phenomenally conscious mental states and therefore there is no justification for the evolution of the mechanisms underlying this new full-blown theory of mind.

Either metacognition is prior or independent of mindreading or HOR theories in general are in trouble. In particular, Carruthers commits himself to the view that phenomenal properties depend on mindreading. As I have tried to show, phenomenally conscious experiences are necessary for being able to ascribe certain mental states to others and therefore prior, at least, to some mind-reading capacities.

MISMATCH PROBLEM AND MISSING MENTAL STATE The final objection I want to consider targets all forms of HOR theories and is related to the problem of misrepresentation between the HOR and the first-order representation and the possibility of a missing first-order state.[46]

What happens if there is no match between the first and the higher-order state? What if the content of the first mental state is RED and I have a higher-order representation to the effect that 'I see GREEN'? What is then the phenomenal character of the experience?[47]

If the reply is as of RED the role of the HOR is unclear, for there is no phenomenological difference between being in a HOR with the content 'I see GREEN' and in another one with the content 'I see RED' as long as the first-order state has the content RED.

Alternatively, holding that in such a case there is no phenomenology seems to be completely ad-hoc. As we learned in the previous chapter, the representational relation has to make room for cases of misrepresentation and this is just a case of misrepresentation.

One could maintain that the content of a higher-order state is some form of indexical content, something like 'I see that'. Although this reply would avoid the problem of a higher-order state misrepresenting the first-order one, it cannot, however, prevent the absence of a first-order state. A problem that we are about to see.

Rosenthal (2005) has maintained that, in the case of a mismatch, the phenomenology is determined by the higher-order thought (as of GREEN). The first-order state plays no role beyond concept acquisition

46 Different versions of this objection have been presented by Block (2011); Neander (1998); Kriegel (2009), etc.

47 Some HOR theories have been developed to avoid this problem (Gennaro (1996); Van Gulick (2004)) in which the first-order state is an essential part of the higher-order state. Without further motivation, however, the reply seems to be ad-hoc unless we are given independent reasons for holding that the first-order mental state is an essential part of the higher-order mental state.

in determining the qualitative character of a conscious experience. In a case in which the subject instantiates a higher-order state but not a first-order one, she is still considered as having a phenomenally conscious experience. But in this situation, there is no state that becomes phenomenally conscious in virtue of being targeted by a higher-order state.

Rosenthal replies that the mental state that is phenomenally conscious is the one that the higher-order thought represents oneself as being in; the conscious mental state is a notional state. In the same sense that something does not need to exist to be the object of my thought, there is no need for the mental state to actually exist to be the target of a higher-order thought. I can have a thought about a pink elephant without there being any pink elephant. Nevertheless, the elephant has the property of being pink. Similarly, in the case of the absent first-order states, non-existent mental states can have the property of being phenomenally conscious.

The problem with this reply is that it is committed to non-existent conscious mental states. If Rosenthal were right there would be phenomenally conscious mental states that have no neural correlate. That seems to me too high a price to pay.

As an alternative to HOR theories, Uriah Kriegel has developed a compelling same-order representational theory of subjective character. This is the focus of the next section.

5.3.2 *Same-Order Representational Theories*

According to same-order representational theories of consciousness, the conscious mental state, M, and the mental state in virtue of which the subject becomes conscious of it, M*, are the same. In this way some of the objections to HOR theories are avoided.

If we are looking for a naturalistic theory of consciousness, and intentionality plays a role in the explanation of the phenomenal character, as it is the case for representationalist theories, we need to provide –or at least give reasons to believe in– a naturalistic theory of intentionality: a theory that explains in natural terms how a certain state can be about another thing. Such theories are on the market (Dretske (1995); Fodor (1990); Millikan (1984)). I have reviewed some of them in the previous chapter, and they seem very promising, though they have their own problems.

The problem that same-order representationalism theory faces is that of explaining in naturalistic terms how a mental state can be about itself in a non-trivial[48] sense that accounts for the subjective character.

Kriegel (2009) makes a very interesting self-representational proposal:

> (Self-representationalism)
>
> For any mental state M of a subject S, M is conscious iff there are M*and $M^{\Diamond}$, such that (i) M*is a proper part of M, (ii) $M^{\Diamond}$ is a proper part of M, (iii) M is a complex of M* and $M^{\Diamond}$, and (iv) M*represents M [indirectly] by representing $M^{\Diamond}$ [directly]. (ibid. 228)

48 There is a trivial sense in which everything is about itself, but this sense obviously does not explain the subjective character.

The idea is that a conscious mental state M is a complex of two parts M^* and $M^\Diamond$ where M^* represents $M^\Diamond$ directly and M indirectly in virtue of representing $M^\Diamond$ and the fact that $M^\Diamond$ is a proper part of M.

In order to evaluate the virtues of (Self-representationalism) we need to first further clarify two notions: that of indirect representation and that of a complex entity.

Kriegel's theory rests on a mereological distinction between sums and complexes (Simons 1987, ch.9). According to Kriegel, M is not a mere mereological sum of $M^\Diamond$ and M^*, but a mereological complex. The difference between mereological sums and complexes is that the way parts are interconnected is not essential for the former but it is for the latter.

> A complex is a whole whose parts are essentially interconnected, or bound in a certain way. A sum is a whole whose parts are interconnected contingently if at all. The interconnection between parts is thus an identity and existence condition of a complex, but not of a sum. (ibid. p.221)

A mereological sum of two elements is the whole that consists of both of them. A complex, contrary to a sum, would cease to exist even if all its parts continued existing. A molecule of H_2O is an example of a complex. For there to be a molecule of H_2O the relation between its parts (H, H, O) is essential: the two atoms of hydrogen have to be covalently bounded to a single oxygen atom.

Furthermore, Kriegel's notion of indirect representation in the case of mental states depends upon something like the following principle:

> If a representation R represents A and A is a part of a complex B, then R represents B indirectly.

Consider the photograph showing the face of my girlfriend I have in my pocket. The photograph directly represents her face and indirectly represents my girlfriend Julia in virtue of the relation between my girlfriend and her face. Suppose that I am a very superficial person and if Julia had a different face I wouldn't date her. Let's call Fjulia to Julia's face and NFJulia to the rest of Julia. Fjulia is a proper part of my current girlfriend. Even if the mereological sum of NFJulia and Fjulia continue existing (because, for instance in a bizarre situation, Julia decided to change her face with John Travolta as in 'Face Off' or because of an strange skin condition her face had to be removed), if Julia doesn't have Fjulia as a face I won't date her. The mereological sum of NFJulia and Fjulia wouldn't be my girlfriend unless the relation between NFJulia and Fjulia holds, namely that Fjulia is Julia's face.[49] My girlfriend is therefore the complex of FJulia and NFJulia, the interconnection between these two parts is an existence condition of my girlfriend as such. In this case, the photo in my pocket represents directly Fjulia and indirectly my current girlfriend because the photo represents Fjulia and Fjulia is a proper part of the complex my girlfriend is. Similarly, in the case of a mental state M of me, M^* represents $M^\Diamond$ directly and M indirectly in virtue of representing $M^\Diamond$ and M^* and $M^\Diamond$being proper parts of the complex M.

49 I am assuming here that 'Julia' is a name for a person and that the person's face is not one of her essential parts. I am not sure about the conditions for personal identity, but I am pretty confident that having the same face is not one of them.

There is a high indetermination in the idea of indirect representation.[50] Imagine that Julia kept her face and we got married, we constitute a complex, I will call this complex 'married couple'. Married couple is not merely the mereological sum of Julia and me: if we get divorced the mereological sum of Julia and me keeps existing but not married couple. It doesn't seem very plausible to hold that the photo my mother took of me after the wedding indirectly represents married couple in virtue of me being a proper part of married couple.

Even if Kriegel can propose a sense of indirect representation that accomodates cases like this, it faces serious problems. Let me present these problems.

Objections to SOR Theories

In what follows I am going to present three objections to same-order representationalism, in particular to Kriegel's proposal. In first place, I will argue that it doesn't seem plausible that the indirect content enters the phenomenology. If this is true, then if for-meness is phenomenologically manifest, then it cannot depend on the indirect content. In second place, I will object that self-representationalism cannot explain the phenomenological observation: what is phenomenologically manifest is that the self is part of the phenomenology, not that the mental state represents itself. Finally, I will argue that it remains unclear what are the conditions for the two parts of the phenomenally conscious state to constitute a complex, a necessary condition for self-representation.

INDIRECT CONTENT AND PHENOMENOLOGY My first objection to self-representationalism is that it is obscure how the indirect content, as Kriegel understands it, can be part of the phenomenology. I am going to present this objection as a dilemma: either there are two contents, in which case the problem of mismatch reappears, or there is just one content, and in this case the indirect content is not part of the phenomenology.

The first horn of the dilemma is to maintain that there are two contents. In the case of the photograph in my pocket, it is about both Fjulia and about my girlfriend: two contents. The problem in such a case is explaining how a single vehicle of representation can have two different contents.

One possibility could be to appeal, for instance, to different consumer systems within a consumer semantics (Carruthers (2000)' theory is an example). In a consumer semantics, the content of a mental state depends, in part, upon the powers of the systems that consume that state. M^* is consumed by two different systems and has therefore two contents (M and $M^{\Diamond}$). This option seems problematic: it is unclear how one could individuate these consumer systems (Millikan (2002)) or how can M^* come to represent M in that case. Be that as it may, the main problem is that the mismatch problem is reintroduced by having two different contents: what happens if M^* has the content M but lacks the content $M^{\Diamond}$?

The second horn is closer to what Kriegel has in mind: there is just one proper content ($M^{\Diamond}$) and a relation of parthood between the directly represented ($M^{\Diamond}$) and the indirectly represented (M). But in this case, I will argue, only the directly represented content enters

50 I am grateful to Manolo Martinez for calling my attention on this fact.

into the phenomenology. This is a problem because indirect content is supposed to explain the subjective character that, for Kriegel, is phenomenologically manifest.

> [Subjective character] is internal to the phenomenology –that is itself a conscious phenomenon. This seems to me self-evident. The very reason to believe in the for-me-ness of experience is fundamentally phenomenological: it is derived not from experiential research, nor from conceptual analysis, nor from any other source, but rather from a certain first-person impression. This suggests that for-me-ness is phenomenologically manifest. Kriegel (ming, p. 4)

For-meness is the property of mental states that determines the subjective character. According to Kriegel, it is the property phenomenally conscious states have of representing themselves. Self-representation is explained through the notion of indirect content; i.e., the mental state indirectly represents itself. This self-representation determines the subjective character of the experience which is phenomenologically manifest. So, if the indirect content doesn't enter the phenomenology, then Kriegel's proposal is jeopardized. Kriegel tries to resist the claim that indirect content is not part of the phenomenology:

> My inclination is to contest the claim that the indirect content of a representation does not show up in the phenomenology [...] one might be tempted to hold that a normal perceptual experience of the sky represents the sky by representing a blue expanse, and yet it seems that both are phenomenologically manifest; or that the olfactory experience of freshly brewed coffee represents the coffee by representing its odor, where again it seems that both are manifest in the phenomenology. However, by the light of the principle that only direct content enters the phenomenology, the sky and the coffee would have to be non-phenomenal. Kriegel (2009, p. 230)

I disagree. It might well be that the coffee is part of the content of the experience, but not part of the content that determines the phenomenal character: the coffee itself is not phenomenologically manifest. What does it mean that the coffee is not phenomenologically manifest? It means that a different substance with the same aroma would give rise to the very same kind of experience.[51] Even if one concedes that these two experiences would differ in content, they do not differ in the content that determines the phenomenal character of the experience, both experiences have the same phenomenal character.

If I smell the aroma of a substance X I have never smelled, seen, nor heard about before, I do not understand how X enters into the experience in the sense of being phenomenally manifest. Consider another substance Y that has the same aroma. The experience I have while smelling X and while smelling Y is exactly the same, consequently neither X nor Y are phenomenologically manifest.[52]

51 I am considering here that two experiences are of the same kind if they have the same phenomenal character.

52 My olfactory experience when I smell the coffee without having any idea what it is and when I know that it is coffee might differ in character. If this were the case, we would have to explain the cognitive penetrability of our phenomenally conscious experiences.

Another example. Imagine someone who has never seen the sky, but has been in an enormous room where the roof is of the same color that the sky. His experience represents a blue expanse. He has never heard about the sky. One day, while he is sleeping we remove the roof of the room. When he wakes up he sees the sky having the very same color as his roof. Now his experience represents a blue expanse, but indirectly represents the sky, because the blue expanse is a proper part of the sky. Nevertheless, the phenomenal character of his experience has not changed at all. One could maintain that the first experience is about the sky and the second about the roof, but this kind of content is irrelevant for the phenomenal character as the example shows.

If this intuition is right, indirect content is not part of the phenomenology and therefore cannot explain the phenomenological observation.

DOES SELF-REPRESENTATIONALISM ACCOUNT FOR FOR-MENESS? My second objection to Kriegel's theory is that self-representationalism fails to account for the subjective character of the experience.

I have argued that a certain form of self-consciousness is a constitutive part of phenomenal consciousness. There are two senses in which self-consciousness can be used as Kriegel (2003, p9. 480-81) has noted. These different uses should be disambiguated. Consider the famous Brentano's claim that every phenomenally conscious state is self-representational. The expression 'M is self-representational' confuses two different uses. The following two quotes illustrate them:

> [Every conscious act] includes within it a consciousness of itself. Therefore, every [conscious] act, no matter how simple, has a double object, a primary and a secondary object, The simplest act, for example the act of hearing, has as its primary object the sound, and for its secondary object, itself, the mental phenomenon in which the sound is heard. Brentano (1973, pp.153-154)

> [T]he mentally active subject has himself as object of a secondary reference regardless of what else he refers to as his primary object. (Brentano (1973, pp. 276-277), also quoted by Kriegel)

There is ambiguity in the use of 'self-representational'. The expression 'M is self-representational' can mean either i) that M represents itself or ii) that M represents the self. I will use 'mental state-involving' to refer to the first use and 'self-involving' to the second one. The first quote by Brentano seems to suggest the first understanding and this one seems to me to be wrong or at least this is not phenomenologically manifest, according to my experience. What my experience reveals is that both the sound and myself are represented by the experience, the former qua object, the latter qua subject. Brentano seems to recognize that in the second quote.

I have suggested that what is phenomenologically manifest is the presence of the qualities of experience for the subject, the phenomenal character is self-involving: what my experience reveals is that both the sound and myself are represented by the experience (the content is self-involving in opposition to merely object-involving).[53] The content of my experience is not merely that such and such is the case, but

53 Kriegel calls self-involving theories egological theories following the phenomenological tradition. Mere object-involving theories are called by opposition non-egological.

that such and such is presented to myself. In phenomenally conscious experiences I do not merely attribute certain properties to the object causing the experience, I attribute to myself being presented with an object with these properties.

Kriegel concedes that the phenomenological observation reveals these facts, but denies that they are constitutive of phenomenal consciousness. What is constitutive of a phenomenally conscious mental state is having a content like 'this mental state is occurring' and not something like 'I am in this mental state'.

> If I were to make another unpedestrian phenomenological assertion, I would say that my current experience's pre-reflective self-consciousness strikes me as egological [self-involving-MS] –that is a form of peripheral self-awareness. My peripheral awareness of my current experience is awareness of it as mine. There is an elusive sense of self-presence or self-manifestation inherent in even a simple conscious experience of the blue sky. It is less clear to me, however, that this feature of peripheral inner awareness [phenomenal character] –its being self-awareness and not mere inner awareness [self-involving and not mere metal state-involving]– is constitutive of the phenomenology.Kriegel (2009, p. 177)

Kriegel holds that whereas the experience is self-involving in normal human adults, this fact is not constitutive of the phenomenology. Pre-reflective self-consciousness is "often egological but not constitutively so." (ibid. p.178). He thinks that infants' and animals' experiences lack this feature. If phenomenal consciousness is essentially self-involving then self-representationalism does not suffice for an experience to have subjective character.

I see no pre-theoretical reason for maintaining that infants' and animals' phenomenally conscious experiences differ in this respect from mine and are not self-involving. It seems to me that a certain form of self is essential for an account of the phenomenal character of experiences: it is phenomenologically manifest that my experiences are somehow experiences of *mine* and not that they represent themselves. Kriegel could claim that my consideration is due to the fact that I am a human adult and human adults' experiences are self-involving. On the other hand infants' or animals' experiences are not self-involving because of the highly cognitive demand that that would require. In the sequel, however, I will offer a notion of self-involving under which it is intuitive that infants and animals may have that kind of states.

Another problem for not self-involving views, as Kriegel's, has to do with my analysis of the content of the experience. The content of the experience is a centered-proposition, a set of worlds and a position within these worlds, a function from centered worlds (world, individual) to truth values. According to mental-state involving (non-egological) views the content is also a kind of centered-proposition, but the proposition is centered not in an individual but in a mental state, it is a function from centered worlds (world, mental state) to truth value: the mental state says about itself that it is occurring. But, if I am right, it also says about the apple that it is disposed to cause an experience as of red in me. A phenomenally conscious mental state M has to represent such a centered feature. Kriegel holds that it also represents itself. Why do we further need this level of representation? As we will see in next

section no further representation of the mental state is required and the neurological structures required for my account are less demanding that those required by Kriegel's self-representationalism.

WHICH IS THE SUFFICIENT CONDITION? Self-representationalism is supposed to account for for-meness, the distinctive property that all and only phenomenally conscious mental states have. For Kriegel's theory it is the property of being a complex of M^* and $M^\Diamond$ where M^* represents $M^\Diamond$. I have serious doubts that this property can guarantee sufficient conditions for being a conscious mental state; in other words, it is not clear that it cannot be satisfied by non-phenomenally conscious mental states.

Metacognition, as we have seen, is the ability to represent our own mental states. We have mental states that are represented by other mental states without thereby giving rise to any phenomenally conscious mental state. If this is true, Kriegel's self-representationalism is jeopardized. Consider a state M_H that represents M_L. Call M_{NC} the aggregate of M_H and M_L, and suppose that M_{NC} is a non-phenomenally conscious mental state. Why is M_{NC} not a phenomenally conscious mental state? The only reply available seems to be that M_{NC}, contrary to M, is not a complex and therefore M_{NC} doesn't represent itself. If we had to appeal to M being phenomenally conscious in order to explain the fact that M is a complex then (Self-representationalism) wouldn't be illuminating at all. So, either there is something in the way that $M*$ and $M^\Diamond$ interact that is different from the way M_H and M_L interact or (Self-representationalism) cannot characterize for-meness.

Kriegel makes a neurological suggestion about the neurological basis of phenomenal consciousness. Although his view on self-representationalism is not committed to this neurological proposal, it can be useful for illustrating my worries. Kriegel locates $M*$'s neural correlate in the dorsolateral prefrontal cortex (dlPFC)[54] and $M^\Diamond$ in the corresponding sensory cortex (the visual cortex for instance in the case of visual experiences). M^* and $M^\Diamond$ are connected via synchronization of their firing rates.[55] Unfortunately for Kriegel connection via synchronization of their firing rates is not exclusive of phenomenally conscious states. If M_H and M_L are connected via synchronization of their firing rates then M_H and M_L are connected by the same way[56] that M^* and $M^\Diamond$. Why M but not M_{NC} is a complex?

According to self-representationalism, a mental state is conscious only if it is a complex that satisfies some further condition (one proper part represents the other) but unless we are given reasons why a phenomenal conscious state like M is a complex and M_{NC} is not, (Self-representationalism) cannot be considered an account of subjective character, for it fails to explain in virtue of what a mental state is a phenomenally conscious mental state.

In the next section I am going to present my own proposal as an attempt to deal with the objections I have presented to other theories

54 In the dream argument on page 191 I showed that the empirical evidence suggests that the dlPFC is not activated during dreams. Kriegel acknowledges that and suggests alternative areas.

55 Kriegel (2009) suggests that synchronization with dlPFC activation is the supervenience base of phenomenal consciousness (ibid. p.280)

56 For instance Cohen et al. (2009) suggest that synchronous neurological oscillations are a plausible mechanism of medial prefrontal cortex-driven cognitive control independent of phenomenal consciousness.

while accounting for the subjective character of the experience. I will call my theory Self-Involving Representationalism (SIR).

5.4 SELF-INVOLVING REPRESENTATIONALISM (SIR)

Phenomenally conscious experiences have phenomenal character, which is constituted by qualitative character and subjective character. All conscious experiences share a subjective character, whereas they differ in qualitative character. Following the taxonomy in 1.3.3, the SIR theory is a compresentist inseparativist view, wherein the subjective character is constitutive of the qualitative character.

The property of having a phenomenally conscious experience is the property of being in a state with certain properties; I have called these properties phenomenal properties. I have argued that phenomenal properties are representational properties of a concrete kind.

When I look at the red apple and when I look at the golf course I undergo two experiences with different phenomenal characters. When I undergo these experiences I am in two different mental states. These two states have different representational content. However, they also have something in common: there is something it is like to be in any of them. For-meness is the one property all and only phenomenally conscious states have –the property of having the right kind of non conceptual *de se* content. Different experiences differ in character in virtue of the differences in concrete *de se* content that the mental states represent.

At the end of the last chapter I dispelled some worries about the circularity of the proposal appealing to the view that I called (Indexical disposition de re). This is not, however, the view we are looking for. According to (Indexical disposition de re), the content of experience is properties of the object causing the experience in me in normal circumstances. As we have seen, this view fails to satisfy certain desiderata for a characterization of the content of experience that determines the phenomenology. Furthermore, it fails to satisfy the phenomenological observation. I have argued that when I have an experience I self-ascribe certain properties to myself; i.e. I ascribe a certain centered feature to the object of the experience (the object that causes the experience in normal circumstances). The content of experience is not merely that such-and-such is the case, but that I am presented with such-and-such. Having an experience, I do not ascribe a property to the object but rather a centered feature. Let me once again remark that ascribing an object with the kind of centered feature we are considering is equivalent to the self-ascription of certain properties. That is to say, the content of the experience is a function from pairs of worlds and individuals to extensions.

Setting the qualitative character aside, I have focused in this chapter on the subjective character of experience and I have reviewed several approaches. In section 5.2, I provided reasons for rejecting theories that try to account for the subjective character of the experience as a form of cognitive access. In section 5.3, I dealt with representational theories of subjective character. These theories claim that a mental state has subjective character if, and only if, the subject is aware of the mental state in *the right way*, where *the right way* is unpacked as a representational relation between mental states (or between proper parts of the mental state).

Explaining the subjective character requires explaining how the content of a mental state can be *for-me*. I have argued that in order to explain the subjective character a certain reference to the self is required. Something can be *for me*, only if there is a me: if there were no me, no self at all, how could something be for me in the relevant sense? How could I self-attribute a certain property to myself? Some representational theories of subjective character avoid a reference to the self. The closest entity to the self that some representational theories present is a subject, where the subject is understood as the holder of mental states (no other commitment is made) and for-meness as a representational relation between mental states. These approaches face serious problems and fail to account for the subjective character of the experience, as I have argued.

In what follows I will maintain the following assumption justified by the phenomenological observation:[57]

> (Self-involving thesis)
>
> Some form of self is required for an explanation of the subjective character of the experience and therefore for any account of the phenomenal character of the experience.

We should clarify what exactly the previous thesis is supposed to entail and how this form of self is supposed to help explain for-meness. These questions are addressed in the remainder of the section, which is organized in five subsections. In 5.4.1 I will clear up the notion of self that is required: the proto-self. Section 5.4.2 will explain how the proto-self helps to account for the differences between states that are phenomenally conscious and those that are not. I will explain, in section 5.4.3, how the SIR theory handles cases of shifted spectrum. In section 5.4.4 I will present my views on the relation between access consciousness and phenomenal consciousness. Some possible objections to SIR will be considered in 5.4.5. Finally I will compare SIR with other competing theories in 5.4.6.

5.4.1 *The Proto-self*

Before getting clear about the notion of self required we should shed light on what the required self is not. The self could merely be the holder of mental states, but for-meness cannot be the property of being held by such a self, for all my mental states are held by the same self and this, *per se*, doesn't give us a distinction between states that are phenomenally conscious and states that are not. We could try to account for the subjective character through a special relation between mental states. This proposal will arguably fail to satisfy the Self-involving thesis, for this sense of self as holder of mental states plays no role unless one of them represents a certain form of self.

On the opposite side, the sense of self involved in pre-reflective self consciousness, the kind of self required for having the required *de se* content, can hardly be something as a narrative self. The narrative or autobiographical self is the elaborate sense of self we usually have in mind when we think about ourselves, with the package of all our

57 In having a phenomenally conscious experience some form of self-consciousness is involved. In the first section of this chapter I tried to illuminate the idea that the kind of self-consciousness required is consciousness of the self as a subject, as opposed to consciousness of one-self as an object.

memories and emotions, aware of the past and anticipating the future. Damasio (2000) presents this notion as follows:

> In complex organisms such as us, equipped with vast memory capacities, the fleeting moments of knowledge in which we discover our existence are facts that can be committed to memory, be properly categorized, and be related to other memories that pertain both to the past and to the anticipated future. The consequence of that complex learning operation is the development of autobiographical memory, an aggregate of dispositional records of who we have been physically and of who we plan to be in the future. We can enlarge this aggregate memory and refashion it as we go through a lifetime. When a certain personal records are made explicit in reconstructed images, as needed, in smaller or greater quantities, they become the *autobiographical self*. (ibid. pp.172-173)

This might be the entity we find in what the phenomenological tradition calls reflective self-consciousness; the entity we recognize ourselves as being. It is plausible that my autobiographical self is what I think about when I think about myself, or what I recognize when I look at myself in the mirror. However, this sense is overly cognitive and it seems completely implausible that this is the sense of self required for subjective character. The autobiographical self requires, as Damasio points out, a complex memory system that is not at all required for phenomenal consciousness. It is highly plausible that human babies and many animals entertain phenomenally conscious states but doubtful that they can have this kind of self-consciousness.

The kind of self should rather be some kind of primitive process that constitutes the basis of such an autobiographical self and lets us explain the subjective character of experience. The required sense of self is the sense of a single, bounded, living organism adapting to the environment to maintain life and the processes that underlie the monitoring of the activity within these bounds. The distinction between what is *me* and what is not has clear evolutionary advantages and allows the evolution of further processes that make use of this representation. The required self is a model of a living body. This is the kind of self I am after. I will call it proto-self in what follows.

One interesting proposal in this direction is Damasio's notion of proto-self. In his book 'The Feeling of What Happens' Damasio (2000) presented a proto-self as a constitutive element of our experiences.[58] I will make use of this very same element as the kind of self involved in the explanation of the subjective character.

According to Damasio,

> The proto-self is a coherent collection of neural patterns which map [represent], moment by moment, the state of a physical structure of the organism in its many dimensions...[t]hese structures are intimately involved in the process of regulating the state of the organism. (Damasio, 2000, p. 154)

58 For a further development of Damasio's ideas about consciousness and the self see Damasio (2010).

> It is an integrated collection of separate neural patterns that map, moment by moment, the most stable aspects of the organism's physical structure.(Damasio, 2010, p. 190)

Damasio presents some of the brain structures that may implement the proto-self. According to Damasio, these structures are necessary for having a phenomenally conscious experience. The proto-self includes ((Damasio, 2000, p. 104)).[59]

- Several brain-stem nuclei: regulate and map body signals. Further independent empirical evidence in favor of the connection between the cortex and the brain-stem has been presented by Churchland (2005); Laureys (2005); Llinas (2002).
- The hypothalamus: maintains a current register of the internal milieu (level of circulating nutrients, concentration of hormones, PH, etc.).
- The insular cortex, the cortices known as S2 and the medial parietal cortices: integrate the representation of the current state of the organism at the level of cerebral hemisphere and the invariant design of the musculoskeletal frame.

The proto-self is a subsystem, composed by a collection of states, that represents my internal states (the internal milieu, viscera, vestibular system and musculoskeletal frame). The proto-self controls and regulates the homeodynamics of the organism, maintaining the required stability for the survival of the organism. These areas do not only monitor but also regulate these internal states; for instance, if the concentration of chemicals in the bloodstream sensed by those areas (brain-stem, hypothalamus, etc) is out of a certain range the system responds to correct this unbalance.[60]

5.4.2 *The Proto-Self and For-meness*

The proto-self is a constitutive element of the *representational* properties that I am identifying with phenomenal properties. Let me elaborate:

My internal states are not phenomenologically manifest in the experience I have while looking at a red apple. The content of the proto-self does not enter the content of phenomenally conscious mental states as such. By that I mean that the proto-self is not an object of the experience as they are the features of the apple. When I was presenting the phenomenological observation, I noted that the experience is not directed to myself as an object but as an experiencing subject; i.e. that I experience my experiences as mine, they are *for-me.* I do not merely attribute certain properties to the object of the experience: I attribute to myself the property of being presented with an entity with certain

59 In his most recent work, Damasio (2010) includes the anterior cingulate cortex as part of the proto-self.

60 Let me remind the reader what the claim that the proto-self represents my internal states amounts to. I will make use of an example for that purpose. There is a little lamp in my TV that indicates whether the TV is on, off or in stand-by. The lamp represents the state of my TV. Properly speaking, there is a system with three possible states: lamp on, lamp off and lamp blinking. Lamp on represents that the TV is on, lamp blinking represents that the TV is in stand-by and lamp off represents that my TV is off. In the same sense, when I claim that the proto-self represents my internal states, what I really mean is that the proto-self is a neurological system composed of different brain states that represent my internal states.

features. This is equivalent to the attribution of the centered features that I have argued constitute the content of phenomenally conscious experiences. When I have a phenomenally conscious experience with phenomenal character PC_{RED} I attribute to the apple the disposition to cause experiences with phenomenal character PC_{RED} *in me*. So, the experience is about the apple and in a sense about myself. I think that this captures the idea in the phenomenological tradition that my experience is directed to the apple as object and to myself as a subject. The proto-self is an element required to have the right kind of de se content.

When I have a visual experience of an apple I do not see anything beyond the apple, the apple is *manifest to me*, and that is what the *for-meness* has to explain. The content of my experience is not merely that such and such is the case, but that such and such is presented to myself. What requires further clarification is the fact that the content of phenomenally conscious experiences is *de se* content.

The content of my experience is a centered feature, a function from worlds centered in me to extensions. In ordinary English, this function can be expressed either by saying that by having an experience as of a red apple I attribute to the apple the centered feature of having the disposition to cause an experience as of a red apple in me (A_{RED}) or by saying that by having an experience as of a red apple I attribute to myself (self-attribute) the property of being confronted with an object that has the disposition to cause experiences as of a red apple in me. The content of this centered proposition is a set of centered worlds, those centered worlds in which the object I am looking at is disposed to cause the experience in me (centered worlds in which I am confronted with the object that causes the experience in normal circumstances).

According to my theory, SIR, differences in the phenomenal character are explained as differences in the representational properties. On the other hand, the subjective character of the experience is determined by a common functional role all and only phenomenally conscious states satisfy. Again, if non-etiological theories of mental content can explain the relation of representation, this common element will be a representational one: phenomenally conscious states have non-conceptual de se representational content.[61]

My next step is to clarify this causal role. There are two elements involved in this explanation: the first one is the proto-self, the second one is what I will call a proto-qualitative state.

On the one hand, the proto-self is a brain structure that has the function of indicating and regulating the homeostasis of the organism. It regulates the internal environment and tends to maintain a stable, constant condition. The theory I am proposing, SIR, holds a the self-involving thesis: it is an egological theory. The kind of self required is the proto-self, a collection of neural patterns that represents and regulates the internal states of the organism.

On the other hand, the proto-qualitative state is a state that has the function of indicating a certain dispositional property. We have seen that the property of undergoing an experience with phenomenal character PC_{RED} is the property of being in a state that represents A_{RED} and

61 As I have already stressed, this leads to an inseparatist proposal where the qualitative character and the subjective character are mutually inseparable. There cannot be an experience with qualitative character without subjective character; there cannot be *redness* without *for-meness*. The subjective character is a constitutive part of the phenomenal character.

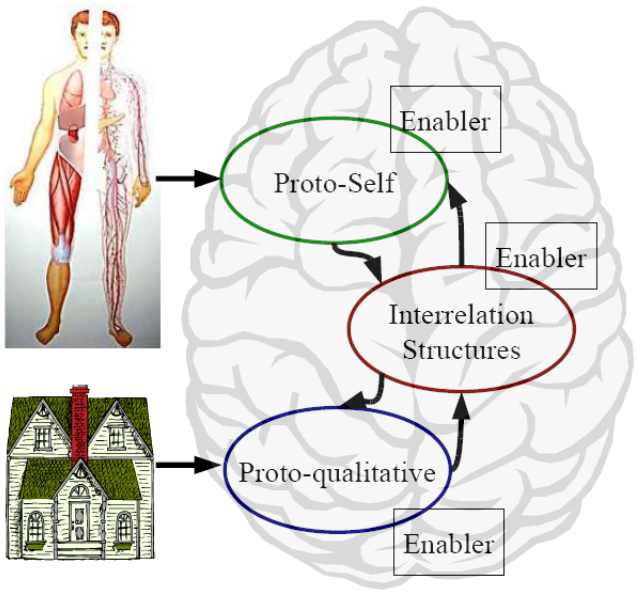

Figure 13: The Proto-Self Interacts with the Proto-Qualitative State.

that an object is A_{RED} only if it is disposed to cause experiences with phenomenal character PC_{RED} in me in normal circumstances. A_{RED} is a centered feature, a function from pairs of worlds and individuals to extensions. In order to get the extension, we need a world and an individual. The proto-qualitative state represents the dispositional property that results from fixing the individual in the previous centered feature.

Different phenomenally conscious states are constituted by different proto-qualitative states. Proto-qualitative states are not phenomenally conscious; i.e. the properties of proto-qualitative states do not suffice for having a phenomenally conscious experience. The proto-self is not a phenomenally conscious state either. It is the interaction between the two that gives rise to a phenomenally conscious mental state which indicates that the property X is affecting the organism. Phenomenally conscious mental states play a differential role in the homeodynamics of the organism. A difference in functional roles accounts for the differences between those mental states that are phenomenally conscious and those that are not (figure 13 illustrates this idea)

At the level of content, this interaction will explain why the content of experience is *de se*. What is relevant for the mental state is not only the properties that the object of the experience (say, the apple) has, that the apple is causing the activation of a certain neural network, but the fact that it is causing the activity of the neural network and that this neural network plays a relevant role in the homeodynamic regulation of a particular organism. The content is not just that the object is disposed to cause such-and-such but that the object is disposed to cause such-and-such in *this* organism, the organism that the proto-self regulates. Other contentful states will also play a role in the organism but not the kind of role that phenomenally conscious states play in homeodynamic regulation.

Let me now present a concrete example to help illuminate the theory.

When looking at the red apple in front of me I undergo a phenomenally conscious experience with phenomenal character PC_{RED}. My visual system will generate a representation of the properties of the apple; this is a proto-qualitative state. Let me refer to this state as PQ_{RED}. Proto-qualitative states are representational states; states that have the function of indicating certain properties. They correspond to the states that (Indexical disposition de re) postulates as phenomenally conscious states, as we saw at the end of the previous chapter.

According to non-etiological theories of function, the function of a trait depends, roughly speaking, on the contribution it makes to the maintenance of the organism it belongs to. PQ_{RED} is a state of my organism, its function is to indicate what produces it via the *particular visual path*$_{PQ_{RED}}$ under *particular lighting conditions*$_{PQ_{RED}}$. An object has the property that PQ_{RED} represents only if the object is disposed to cause the activation of PQ_{RED} in an organism like mine[62] via the *particular visual path*$_{PQ_{RED}}$ under *particular lighting conditions*$_{PQ_{RED}}$. If an object reflects light with a wavelength of 650nm in these lighting conditions, then it can cause PQ_{RED} via the *particular visual path*$_{PQ_{RED}}$. The surface of the apple reflects light, in these lighting conditions, with a wavelength of 650 nm and is therefore represented by PQ_{RED}.

At the end of the previous chapter we saw that, under the plausible assumption that necessarily coextensive properties are identical, the property of having the disposition to cause PQ_{RED} in normal conditions is identical to the property of emiting light with a wavelenght of 650nm or reflecting light with this wavelenght under certain lighting conditions (the disjunction of the categorical bases of the disposition to cause the state in normal conditions) and therefore there is nothing circular in this view.

The proto-qualitative state has the function of indicating the property of emitting light with a wavelength of 650nm or reflecting light with this wavelength under certain lighting conditions, but this is, still, an unconscious representation; *the content of the experience is not this property*.

On the other hand, I have a representation of my internal states: the proto-self. This latter representation is altered by the processing of the apple (change in the retina or in the muscles that control the position of the eyeball, but also changes in the smooth musculature of the viscera, at various places of the body, corresponding to emotional responses, some of them innate). The interaction between the proto-qualitative state and the proto-self constitutes a mental state with the content 'redness for-me', a conscious mental state. When my organism is in this state, it attributes to itself the property of being presented with an object that is disposed to cause PQ_{RED} in *this organism* in normal circumstances.[63] This is equivalent to the claim that my organism attributes to the apple the centered feature A_{RED}. If the object causing the experience has the disposition to cause it in normal circumstances (if the object causing the experience has the property of emitting light with a wavelength of 650nm or reflecting light with this wavelength under certain lighting

62 As we have seen, according to the organizational account, two organisms are alike if and only if they have the same organization (See 4.4.2).

63 This does not prevent that another organism can attribute to itself the very same property, just as Manolo and Adrian attribute to themselves the very same property when they have the belief expressed by the sentence 'I live in Barcelona'.

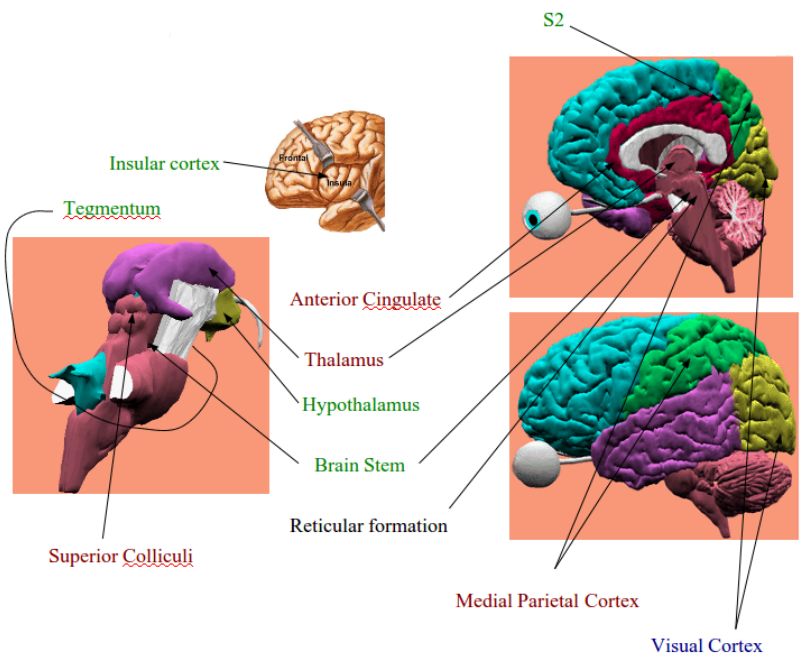

Figure 14: Structures Involved in Phenomenal Consciousness

conditions) then my experience is correct or veridical, otherwise it is not.

The conscious mental state is the complex formed by the proto-self, the proto-qualitative state and their relation. Contrary to Kriegel's proposal, I see no particular problem in naturalizing the content of this state.[64]

Having such an experience is being in a mental state with certain properties that I have called phenomenal properties; they determine the phenomenal character of the experience. The Total Neural Basis (TNB) of an experience with phenomenal character PC_{RED} is the neural activity minimally sufficient for having an experience with phenomenal character PC_{RED}, we can call it TNB_{RED}. The Core Neural Basis (CNB) is the part of the TNB that distinguishes mental states with PC_{RED} from other phenomenally conscious states. This core neural basis corresponds to what I have called proto-qualitative state. According to the SIR theory, TNB_{RED} is constituted by the proto-qualitative state (CNB), the proto-self and the structures that implement the interaction between the proto-self and the proto-qualitative state plus the mechanisms that allow these areas to perform its function.[65] Figure 14 illustrates some of the involved areas, according to the colors in figure 13.

As we saw in the previous chapter, we do not want to maintain that all the properties of TNB_{RED} are necessary for phenomenal consciousness. SIR helps selecting some of the TNB_{RED} properties as relevant for having the experience. According to the SIR theory, phenomenal properties are identical to the set of properties necessary for having the causal role in virtue of which TNB_{RED} represents A_{RED}, the disposi-

64 Further work has to be done for this purpose, but the proposal shows the avenue for naturalizing this de se content by appealing to the role the state plays in the stability of the system.

65 I call these mechanisms enablers. An example of an enabler is the reticular formation.

tion to cause experiences with phenomenal character PC_{RED} in me in normal circumstances.

SIR theory is a first order theory, there is no need for a further representational state for it to qualify as phenomenally conscious. We can see this more explicitly if we compare it with what I consider to be a fair reading of Damasio (2000)'s proposal. According to Damasio, the relation between the proto-self and what I have called proto-qualitative state has to be represented by higher-order structures. The areas responsible for this representation include: the cingulate cortex, the thalamus, the superior colliculi and some pre-frontal cortices.[66] These structures must be capable of exerting an influence on the first-order representations. According to SIR theory these structures realize the interaction between the proto-self and the proto-qualitative state.

Damasio maintains that these higher-order structures map the relation between what I have called proto-qualitative state and the proto-self. Damasio's theory is a HOR theory. I have already presented the problems with HOR theories in the previous section. Damasio postulates the role of those structures due to i) their relation to consciousness and ii) the need for mapping the changes in the proto-self.

The problem with this reading is that it reintroduces the problem of mismatch and the higher-order representation seems not to be sufficient for phenomenal consciousness. If the higher-order structures were contentful states, then, an independent activity of the higher-order structures (without the first-order one) corresponding to an experience of a red apple, would produce an experience as of a red apple. But there are evidences showing that the brain-stem nuclei, which are part of the first-order structures, are necessary for consciousness.

For a higher-order approach the second-order states are about the relation between the proto-self and the proto-qualitative state. For the SIR theory, on the other hand, these structures are not contentful states

66 Damasio provides some support for his claim that these structures are required:
The cingulate cortex comprises a combination of sensory and motor roles and it is involved in a large variety of complex movements including those of the viscera. Lesion and fMRI studies relate this area with emotion, attention and autonomic control. The evidence presented by Damasio is based on the reduction of the activity in this area on slow-wave sleep (compared to a significant increment during REM), hypnosis and some forms of anesthesia. Bilateral anterior lesion of the cingulate causes a condition known as akinetic mutism, that is described by Damasio as "suspended animation, internally as well as externally"(Damasio, 2000, p. 176). In his most recent work Damasio (Damasio, 2010) includes the cingulate cortex as part of the proto-self, however akinetic mutism is usually characterized as a variant of minimally conscious states. The relation between the anterior cingulate cortex and consciousness is nevertheless a controversial one:

> The interpretation is complicated by the fact that, in the rare instance in which such patients recover, there is usually amnesia for the akinetic episode, as in the original case of Cairns, though one patient who eventually recovered reported that she remembered the questions posed by the doctor but did not see a reason to respond (Laureys and Tononi (2008, p. 395))

Patients with bilateral medial parietal damage, in spite of being awake, "...do not look at anything with any semblance of attention, and their eyes may stare vacantly or oriented toward objects with no discernible motive." (Damasio (2000, p. 178)) This lesion is also found in patients with Alzheimer disease.
The superior colliculi receives a multiplicity of sensory inputs from several modalities and communicates the results to a variety of brain stem nuclei, the thalamus and the cortex. Damasio recognizes there is no evidence in humans that the superior colliculi supports consciousness in the absence of thalamic and cingulate structures, even assuming that the brain-stem structures remain intact.
According to Damasio, the idea that the thalamus is related to consciousness is mainly based "on credible experiments in animals and on the likelihood that abnormal discharges in absence seizures, during which consciousness is disrupted, originate in the thalamus" (ibid. p.178)

or, more precisely, states with a content that is relevant for phenomenal consciousness; they merely implement the relation between the proto-self and the proto-qualitative state. The higher-order approach is committed to the view that there might be phenomenal consciousness without the proto-self or the proto-qualitative state, a case of misrepresentation. On the other hand, first-order theories maintain that the activation of the proto-qualitative state and the proto-self are necessary for phenomenal consciousness.

We should not postulate further representations when they are not explanatorily required, and in this case they are not needed. Second-order structures, if needed for consciousness, are not contentful states; some of them could be a constitutive part of the conscious mental states because they implement the relation between the proto-self and the proto-qualitative state.

In (Damasio, 2010), Damasio seems to have changed his mind. He does not refer to these structures as higher-order maps but as neural coordinators of the relation between the proto-self and what I have called proto-qualitative states:

> [T]he modified protoself *must be connected* with the images of the causative object [proto-qualitative state]...Might there be a need for neural coordinating devices to create the coherent narrative that defines the protoself? The answer depends on how complex the scene is and whether it involves multiple objects... There are good candidates for that role, located at the subcortical level. The first candidate is the superior colliculus...The second candidate for the role of coordinator is the thalamus. (ibid. p.460, my emphasis)

The "second-order" structures should not be considered second-order in any sense but same-order structures.

5.4.3 *SIR and the Shifted Spectrum*

In this subsection I will expound the way SIR handles cases of shifted spectrum. This will further help to clarify the theory.

Consider my red apple. In normal lighting conditions the apple reflects light with a wavelength of, say, 650nm (the apple would not reflect light with a unique wavelength but we can, I think, abstract from this problem).

When Marta and I look at the apple we undergo slightly different experiences. Marta's visual system and mine are slightly different and we undergo slightly different experiences when we look at the very same apple under the very same lighting conditions; both of these experiences are correct. We can call the phenomenal character of these two experiences PC_{RED1} and PC_{RED2} respectively.

When Marta and I undergo these experiences, we are in different proto-qualitative states, PQ_2 and PQ_1 respectively. PQ_1, a state of my organism, has the function of indicating what produces it via the $particular\,visual\,path_{PQ1}$ under $particular\,lighting\,conditions_{PQ1}$. An object has the property that PQ_1 represents only if the object is disposed to cause the activation of PQ_1 in an organism like mine via the $particular\,visual\,path_{PQ1}$ under $particular\,lighting\,conditions_{PQ1}$. If an object reflects light with a wavelength of 650nm in these lighting conditions, then it can cause PQ_1 via the $particular\,visual\,path_{PQ1}$.

The surface of the apple reflects light, in these lighting conditions, with a wavelength of 650 nm and is therefore represented by PQ_1.

PQ_2 is a state of Marta's organism, the function of this state is to indicate what produces it via a $particular\,visual\,path_{PQ2}$ under $particular\,lighting\,conditions_{PQ2}$. An object has the property that PQ_2 represents only if the object is disposed to cause the activation of PQ_2 in an organism like Marta's via the $particular\,visual\,path_{PQ2}$ under $particular\,lighting\,conditions_{PQ2}$. As in my case, if an object reflects light with a wavelength of 650nm, in these lighting conditions, then it can cause PQ_2 via the $particular\,visual\,path_{PQ2}$. If an object reflects light with a wavelength of 650nm in these lighting conditions, then it can cause PQ_2 via the $particular\,visual\,path_{PQ2}$. The surface of the apple reflects light with a wavelength of 650 nm in these lighting conditions, and is therefore represented by PQ_2. Our proto-qualitative states PQ_1 and PQ_2 have the very same content: they both represent objects that emit or reflect light with a waveleght of 650 nm.[67]

PQ_1 interacts with my proto-self, the system that monitors and controls the homeodynamics of my organism. The state that results from this interaction is a *phenomenally conscious mental state*. This state represents that the organism is presented with an object that is disposed to cause PQ_1 in normal conditions (*via* $particular\,visual\,path_{PQ1}$ under $particular\,lighting\,conditions_{PQ1}$). When the organism is in this state it attributes to itself the property of being presented with an object that is disposed to cause PQ_1 in normal conditions: it attributes to the object a centered feature (A_{RED_1}).

Marta, on the other hand, attributes to herself the property of being presented with an object that is disposed to cause PQ_2 in normal conditions. Marta attributes to the apple A_{RED_2}. Marta's experience and mine differ in character because we are self-attributing different properties; that is to say we are attributing to the apple different centered features. Both experiences are, nevertheless, correct.

Phenomenal properties, the properties that my mental state has such that when I am in this state I undergo a phenomenally conscious experience, are representational properties whose content is *de se*.

Imagine that Marta can be in a proto-qualitative state PQ_1 whose function is to indicate what produces it via the $particular\,visual\,path_{PQ1}$ under $particular\,lighting\,conditions_{PQ1}$. In the case of Marta, an object can cause PQ_1 via $particular\,visual\,path_{PQ1}$ if the object can reflect light under $particular\,lighting\,conditions_{PQ1}$ with a wavelength of 654nm. The fire engine on the corner of her street can reflect light with a wavelength of 654nm under $particular\,lighting\,conditions_{PQ1}$ and therefore is represented by PQ_1. Imagine further that this state interacts with her proto-self, the corresponding phenomenally conscious state represents than the organism is presented with an object that is disposed to cause PQ_1 in normal conditions. When she looks at the fire engine she undergoes a phenomenally conscious experience with the very same character as the one I undergo when I look at the red apple. I attribute the same centered feature to the apple that Marta attributes to the fire engine; we are both attributing to ourselves the same property. Both attributions are correct.

67 I am assuming here for simplicity that the lighting conditions the function of PQ_1 and PQ_2 determine are the same, so that when we both look at the apple under the same lighting conditions we can both have veridical experiences.

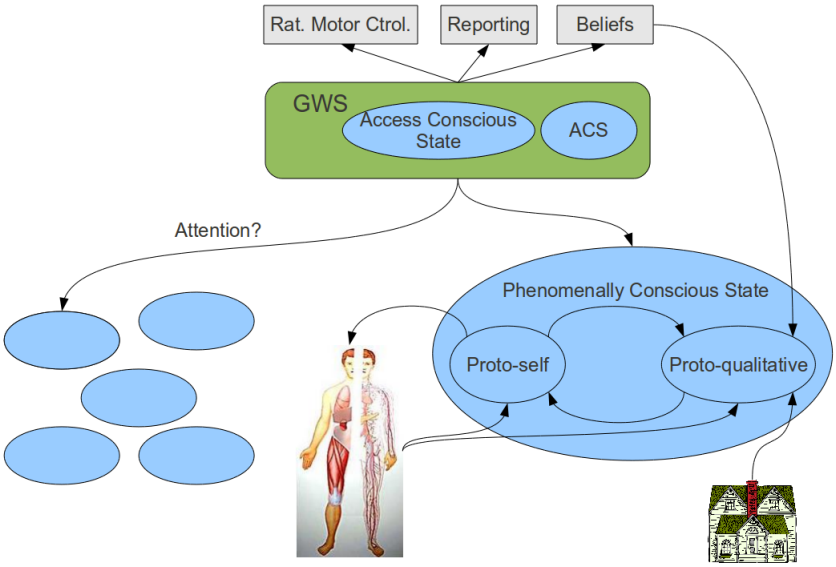

Figure 15: Access Consciousness in the SIR Theory

5.4.4 *SIR and Access Consciousness*

In 5.2 I offered some arguments for rejecting the view that phenomenally conscious states are always accessible for report, belief-forming and rational motor control. In this section I will elaborate on the relation between access and phenomenal consciousness, according to SIR theory.

In the introduction I presented an interesting proposal, made by Uriah Kriegel (2009), about the relation between phenomenal consciousness and access consciousness: phenomenal consciousness is the categorical basis of access consciousness. I think that phenomenal consciousness is part of this categorical basis but not the whole story. Figure 15 illustrates the idea.

Phenomenally conscious states are states constituted by the relation of a proto-qualitative state and the proto-self. The proto-self receives inputs from the internal states and controls their stability. The proto-qualitative state represents certain properties of the environment or the body (in the case of emotion, pain, orgasm, etc.). The interaction between them gives rise to a phenomenally conscious state, i.e. a state with the right kind of de se content.

In the figure, blue ellipses represent phenomenally conscious states. These states constitute the iconic memory postulated by Sperling (see p. 188). On the other hand, we may assume that being encoded in the GWS is the categorical basis of access consciousness (described at the neurological level). States there are freely available for report, rational control of action and belief forming are states encoded in the GWS. Being a phenomenally conscious state does not suffice for gaining access to the GWS; there are phenomenally conscious states that are not encoded there.

Recall that, according to the GWS theory, allied processes compete for access to the global workspace, striving to disseminate their messages to all other processes in an effort to recruit more cohorts and thereby

increase the likelihood of achieving their goals. Phenomenally conscious mental states have good chances of gaining access to it. The proto-qualitative state and the proto-self are examples of those assemblies. The recurrent loops between them that help to constitute the phenomenally conscious mental state increase the likelihood that the phenomenally conscious state will access the GWS. Only some of these states gain access to the GWS and arguably further processes are required in order to gain access to the global workspace. Attention is likely one of the mechanisms involved.[68]

One point that should be made is that phenomenal consciousness seems to be cognitively penetrable. In an example on page 213 I suggest that the phenomenal character of an olfactory experience may vary depending on whether or not I believe that what I am smelling is coffee. I predict that this cognitive influence is reflected in the proto-qualitative state or alternatively at the level of the intermediate structures. In this case, we can think of these intermediate structures as a channel that modulates the communication between the proto-self and the proto-qualitative state without having to postulate any further representation.

In the next subsection I will consider some possible objections to SIR theory and provide a rejoinder.

5.4.5 *Objections to SIR and Rejoinders*

One possible objection to my proposal is that the role of the proto-self structures in consciousness can be seen as causal and not constitutive.[69] One way of presenting this objection is by asking whether someone whose blood chemicals were regulated by an external computer that did not interact with her mental representations would be phenomenally conscious. The proto-self would not be the system performing the homeodynamic control of the organism, but an external computer. Proto-qualitative states would not interact with this external computer and therefore they would not interact with the system responsible for the homeodynamic control of the organism. This seems to suggest that there would not be states that satisfy the relevant functional role and therefore, according to the SIR theory, this hypothetical subject would not undergo phenomenally conscious experiences. I think, however, that she would.

Having a phenomenally conscious experience is being in a phenomenally conscious mental state. A phenomenally conscious mental state is a brain state with certain representational properties –a brain state that satisfies a certain functional role.

In the example above, the homeodynamics of the subject are regulated by an external device, not by the proto-self. However, the proto-self still has the function of indicating and regulating the homeodynamics of the organism; a trait has a function independently of whether it

68 On page 126, I presented Block's objection to representationalism based on Carrasco's experiment. He maintains that attention modifies the phenomenal character of the experience. As I have argued, it is not at all clear that Carrasco's experiments support this claim. Furthermore, attention is not a unique mechanism (Kastner (2010)), so even if some of these mechanisms modify the phenomenal character I predict that this happens at the level of the proto-qualitative state and that a different mechanism is involved in accessing the GWS.
See Prinz (ming) for a theory of consciousness that maintains that phenomenally conscious states are attended representational states and an interesting discussion of empirical evidence about the role of attention in consciousness.

69 I am grateful to Ned Block for raising this objection.

actually performs it or not. For instance, in the organizational account presented in subsection 4.4.2 on page 161, the proto-self is produced and maintained for indicating and regulating the homeodynamics of the organism and therefore this is its function whether it does so or not. The proto-self is malfunctioning, but the phenomenal character of the experience depends on its function, not on its actually performing such a function. If there is an interaction between the proto-qualitative states and the proto-self then the subject in the mental thought experiment will undergo phenomenally conscious experiences.

One could also object that the proposal seems to be 'too contingent':[70] we can imagine spiritual entities that are conscious despite lacking a proto-self. Undoubtedly, such spiritual entities seem to be conceivable, but maybe on reflection they are not. The phenomenal character of the experience is explained as a complex intentional content 'X for-me'. In order to have this content a proto-self is required. Maybe Damasio's proposal is not right or it is not the only possible way of having a proto-self, but if spiritual entities lack a proto-self then they cannot be conscious, according to SIR; they cannot have states with the right kind of de se content. Furthermore, we learned in the second chapter that materialists that endorse the phenomenal concept strategy should resist the entailment between conceivability and possibility. Phenomenally conscious spiritual entities are conceivable but not metaphysically possible unless their mental states have for-meness.

A third possible objection is that my theory is committed to the non-existence of *neutral* mental states that are conscious. A mental state is *neutral* if the instantiation of this mental state by the subject does not modify her proto-self. What happens if the modification of the proto-self happens spontaneously? What happens if a subject instantiates a proto-qualitative state that is not accompanied by the corresponding modification of the proto-self? The reply to these questions cannot be based on the idea that there cannot be a proto-qualitative state without a modification of the proto-self or the other way around. Surely the modification of the proto-self corresponding to my visual experience of a red patch is accompanied by the corresponding proto-qualitative state and in normal circumstances a proto-qualitative state is accompanied by a certain modification of the proto-self. But both can happen independently of each other.[71]

The answer is that in both circumstances we do not have a conscious mental state. The whole complex constitutes a conscious mental state, because both, the proto-self and the proto-qualitative state, are required to have a phenomenally conscious mental state, a mental state with the proper content; namely, a state with the proper causal role. If one of their parts or the relation between them is missing we have no complex. Both parts are required to fulfill a certain causal role, the role that allow the traits to have the function of indicating a certain centered feature.

To sum up, let me repeat that, in our case, a phenomenally conscious state is a brain state, concretely TNB. The phenomenal character of the experience is determined by certain properties of the TNB, the properties in virtue of which TNB represents a certain centered feature. In the case of a phenomenally conscious experience with phenomenal character PC_{RED}, if TNB_{RED} is its total neural basis, then phenomenal

70 I am grateful to Uriah Kriegel for pressing me on this issue.

71 The kind of possibility involved in posing this problem is simply a metaphysical possibility. Not even nomological possibility is required, although I see no reason to deny the latter.

properties are the properties in virtue of which TNB_{RED} represents A_{RED}. I have argued that TNB_{RED} includes two parts, the proto-self and the proto-qualitative state properly related to each other.

It has to be noted that the proto-self and the proto-qualitative state are theoretical postulations that would determine the phenomenal character of the experience. Damasio's version of the proto-self is a very plausible candidate for the proto-self role required by the theory. It is a matter of future empirical research to look for structures with the causal role postulated by the theory. If such structures are not found in the brain the theory would turn out to be false.

5.4.6 *Comparison of SIR with Competing Theories*

Before closing the chapter, I would like to compare the SIR theory with other competing theories and show how it solves many of the objections presented against them.

- According to SIR, phenomenal properties are intrinsic properties of the subject. This saves the intuition that microphysical duplicates are phenomenological duplicates contrary to externalist representational theories.

- SIR accounts for the content of the experience. The content of a phenomenally conscious experience is de se, something like X has the disposition of causing TNBred *in me*. It is not merely that it has the disposition of causing TNBred but TNBred *in me*. The proto-self is a constitutive part of TNBred.

- For SIR, the conditions for a mental state to be phenomenally conscious do not depend on the subject's cognitive capacities. A theory of consciousness that does not depend on the subjective cognitive capacities matches the empirical evidence better (mess argument). Furthermore, it does not essentially involve the cognitive mechanisms underlying reportability, so it doesn't face the objection based on the empirical evidence from dream experiences (dream argument). SIR can help to explain the cognitive access we have to our phenomenally conscious mental states.

- Contrary to theories that depend on high cognitive abilities, SIR theory is not too demanding. There is no problem in ascribing phenomenal consciousness to infants and higher animals.

- SIR is not a higher-order theory. Consequently, it faces neither the problems derived from a possible mismatch between the first-order and the higher-order mental state (mismatch objection) nor the problems derived from the missing first-order mental state (missing first-order objection). A phenomenally conscious mental state is a mental state that has certain content, for the mental state to have such content both the proto-self and the proto-qualitative state are necessary. If one of these states is missing –namely, if there is no proto-qualitative state or there is no proto-self involved– then there is no phenomenally conscious state. There is no state that satisfies the required functional role.

- According to SIR, phenomenal consciousness doesn't depend on a Theory of Mind faculty, nor does it depend on metacognition or

mindreading. The truth of SIR doesn't depend on the result of the discussion about the priority between mind-reading and metacognition. As we have seen, this poses a problem for Carruthers' theory and jeopardizes other theories.

- SIR theory is a self-involving theory. This matches the phenomenological observation.
- A phenomenally conscious mental state is a mental state we are Aware of in virtue of its having for-meness. It is not true, according to SIR, that a mental state M has for-meness because we are Aware of it. We are Aware of M in virtue of its causal role (for-meness).

SIR theory is a compelling theory. It accounts for the phenomenological observation and explains the content of experience while solving many of the problems of competing theories.

In this work I have attempted to present the mainstays of the SIR theory. Further philosophical and empirical research is required to elaborate the details of the theory and for evaluating the merits of the proposal.

BIBLIOGRAPHY

Amstrong, D.: 1968, *A Materialist Theory of the Mind*. London: Routledge. (Cited on pages 201 and 203.)

Anton-Erxleben, K., C. Henrich, and S. Treue: 2007, 'Attention changes perceived size of moving visual patterns'. *Journal of Vision* **7**(11). (Cited on page 125.)

Antony, M., 'Are Our Concepts Conscious State and Conscious Creature Vague?'. (Cited on page 101.)

Antony, M.: 2006a, 'Consciousness and Vagueness'. *Philosophical Studies* **128**(3), 515–538. (Cited on pages 94, 100, 101, and 102.)

Antony, M. V.: 2006b, 'Papineau on the vagueness of phenomenal concepts'. *Dialectica* **60**(4), 475–483. (Cited on page 97.)

Armstrong, D. M.: 1981, 'What is consciousness?'. In: *The Nature of Mind*. Cornell University Press. (Cited on pages 13 and 23.)

Artiga, M., 'Re-Organizing Organizational Accounts of Function'. (Cited on page 162.)

Ayala, F. J.: 1970, 'Teleological explanations in evolutionary biology'. *Philosophy of Science* **37**(1), 1–15. (Cited on page 148.)

Aydede, M. and G. Guzeldere: 2005, 'Cognitive Architecture, Concepts, and Introspection: An Information-Theoretic Solution to the Problem of Phenomenal Consciousness'. *Noûs* **39**(2), 197255. (Cited on page 66.)

Baars, B.: 2009, 'Global Workspace theory'. In: A. C. Tim Bayne and P. Wilken (eds.): *The Oxford Companion to Consciousness*. Oxford university Press. (Cited on page 185.)

Baars, B. J.: 1988, *A Cognitive Theory of Consciousness*. Cambridge University Press. (Cited on pages 184 and 185.)

Baker, L.: 1937, 'The influence of subliminal stimuli on verbal behaviour'. *Journal of experimental psychology* **20**, 84–100. (Cited on page 179.)

Balog, K.: 1999, 'Conceivability, possibility, and the mind-body problem'. *Philosophical Review* **108**(4), 497–528. (Cited on pages 49 and 75.)

Balog, K.: 2009, 'Phenomenal Concepts'. In: B. McLaughlin, A. Beckermann, and S. Walter (eds.): *Oxford Handbook in the Philosophy of Mind*. Oxford University Press. (Cited on page 66.)

Balog, K.: forthcominga, 'Acquaintance and the Mind-Body Problem'. In: C. Hill and S. Gozzano (eds.): *The Mental, the Physical*. Cambridge University Press. (Cited on page 80.)

Balog, K.: forthcomingb, 'Illuminati, zombies and metaphysical gridlock'. (Cited on pages 50 and 52.)

Balog, K.: forthcomingc, 'In Defense of the Phenomenal Concept Strategy'. *Philosophy and Phenomenological Research*. (Cited on pages 66, 76, 79, and 81.)

Barandiaran, X. and A. Moreno: 2008, 'Adaptivity: From Metabolism to Behavior'. *Adaptive Behavior* **16**, 324–324. (Cited on page 162.)

Barker-Plummer, D.: 2011, 'Turing Machines'. http://plato.stanford.edu/entries/turing-machine/. (Cited on page 36.)

Bermudez, J. L.: 1998, *The Paradox of Self-Consciousness*. The MIT Press. (Cited on page 175.)

Bermudez, J. L.: 2004, 'Vagueness, phenomenal concepts and mind-brain identity'. *Analysis* **64**(2), 131–139. (Cited on page 97.)

Block, N.: 1978, 'Troubles with functionalism'. *Minnesota Studies in the Philosophy of Science* **9**, 261–325. (Cited on pages 37 and 38.)

Block, N.: 1986, 'Advertisement for a semantics for psychology'. *Midwest Studies in Philosophy* **10**(1 ,), 615–78. (Cited on page 205.)

Block, N.: 1990, 'Inverted Earth'. *Philosophical Perspectives* **4**(4), 53–79. (Cited on pages 120 and 132.)

Block, N.: 1995-2002b, 'On a confusion about the function of consciousness'. In: N. Block (ed.): *Consciousness, Function, and Representation: Collected Papers*, Vol. 1. Bradford Books. (Cited on page 12.)

Block, N.: 1996, 'Mental paint and mental latex'. *Philosophical Issues* **7**, 19–49. (Cited on page 124.)

Block, N.: 2002a, 'The Harder Problem of Consciousness'. *Journal of Philosophy* **99**(8), 391–425. (Cited on pages 41 and 96.)

Block, N.: 2002c, 'Some concepts of consciousness'. In: D. Chalmers (ed.): *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press. (Cited on page 12.)

Block, N.: 2003, 'Mental paint'. In: M. Hahn and B. Ramberg (eds.): *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Mit Press, pp. 165–200. (Cited on pages 119, 120, 121, 127, and 128.)

Block, N.: 2006, 'Max Black's objection to mind-body identity'. *Oxford Review of Metaphysics* **3**. (Cited on pages 66 and 80.)

Block, N.: 2007a, 'Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience'. *Behavioral and Brain Sciences* **30**, 481–548. (Cited on pages 23, 183, 188, and 191.)

Block, N.: 2007b, 'Is experiencing just representing?'. In: *Consciousness, Function and Representation*. The MIT Press. (Cited on pages 153 and 155.)

Block, N.: 2007c, 'Wittgenstein and Qualia'. *Philosophical Perspectives* **21**(1), 73–115. (Cited on page 127.)

Block, N.: 2009, 'Comparing the Major Theories of Consciousness'. In: M. Gazzaniga (ed.): *The Cognitive Neurosciences IV*. MIT Press. (Cited on page 186.)

Block, N.: 2010, 'Attention and mental paint'. *Philosophical Issues* **20**(1), 23–63. (Cited on pages 125 and 126.)

Block, N.: 2011, 'The Higher Order Approach to Consciousness is Defunct'. *Analysis* **71**(3). (Cited on page 209.)

Block, N.: forthcoming, 'Functional reduction'. In: D. S. Terry Horgan and M. Sabates. (eds.): *Supervenience in Mind: A Festschrift for Jaegwon Kim*. (Cited on page 43.)

Block, N. and J. A. Fodor: 1972, 'What psychological states are not'. *Philosophical Review* **81**(April), 159–81. (Cited on page 37.)

Block, N. and R. Stalnaker: 1999, 'Conceptual analysis, dualism, and the explanatory gap'. *Philosophical Review* **108**(1), 1–46. (Cited on pages 61, 62, and 63.)

Boghossian, P. A. and J. D. Velleman: 1989, 'Color as a secondary quality'. *Mind* **98**(January), 81–103. (Cited on page 134.)

Boorse, C.: 2002, 'A Rebuttal on Functions'. In: R. C. A. Ariew and M. Perlman (eds.): *Function*. (Cited on page 159.)

Boyd, R.: 1980, 'Materialism Without Reductionism: What Physicalism Does Not Entail'. In: N. Block (ed.): *Readings in the Philosophy of Psychology*, Vol. 1. Cambridge University Press. (Cited on page 35.)

Braun, A., Balkin, T.J., Wesenten, N.J., R. Carson, M. Varga, P. Baldwin, S. Selbie, Belenky, G., and P. Herscovitch: 1997, 'Regional cerebral blood flow throughout the sleep wake cycle. An H2(15)O PET study'. *Brain* **120**, 1173–1197. (Cited on page 195.)

Brentano, F.: 1874/1973, *Psychology from an Empirical Standpoint*. International Library of Philosophy. (Cited on pages 201 and 214.)

Brogaard, B., 'Degrees of Consciousness'. (Cited on page 98.)

Brook, A.: 2008, 'Kant's View of the Mind and Consciousness of Self'. http://plato.stanford.edu/entries/kant-mind/. (Cited on page 177.)

Brown, R.: 2010, 'Deprioritizing the A Priori Arguments against Physicalism'. *Journal of Consciousness Studies* **17**(3-4). (Cited on pages 52, 53, and 57.)

Bubl, E., E. Kern, D. Ebert, M. Bach, and L. T. van Elst: 2010, 'Seeing Gray When Feeling Blue? Depression Can Be Measured in the Eye of the Diseased'. *Biological Psychiatry* **68**(2), 205–208. Vascular Function in Depression in Older Adults. (Cited on page 19.)

Burge, T.: 1979, 'Individualism and the mental'. *Midwest Studies in Philosophy* **4**(1 ,), 73–122. (Cited on pages 69 and 131.)

Burge, T.: 1997, 'Two Kinds of Consciousness'. In: O. F. Ned Block and G. Guzeldere (eds.): *The Nature of Consciousness: Philosophical Debates*. Cambridge: MIT Press. (Cited on pages 119 and 184.)

Burge, T.: 2007, *Foundations of Mind (Philosophical Essays)*. Oxford University Press, USA. (Cited on pages 174 and 201.)

Byrne, A.: 2001, 'Intentionalism Defended'. *Philosophical Review* **110**(2), 199–240. (Cited on pages 87, 122, 123, and 124.)

Carrasco, M.: 2006, 'Covert attention increases contrast sensitivity: Psychophysical, neurophysiological and neuroimaging studies'. *Progress in Brain Research* **154**, 33–70. PMID: 17010702. (Cited on page 125.)

Carruthers, P.: 2000, *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge University Press, 1St edition edition. (Cited on pages 23, 201, 203, 204, 205, 207, and 212.)

Carruthers, P.: 2003, 'Phenomenal concepts and higher-order experiences'. *Philosophy and Phenomenological Research* **68**(2), 316–336. (Cited on page 66.)

Carruthers, P.: 2009, 'How we know our own minds: The relationship between mindreading and metacognition'. *Behavioral and Brain Sciences* **32**(2), 121–138. (Cited on pages 207 and 208.)

Carruthers, P. and B. Veillet: 2007, 'The phenomenal concept strategy'. *Journal of Consciousness Studies* **14**(9-10), 212–236. (Cited on pages 76 and 77.)

Caston, V.: 2002, 'Aristotle on consciousness'. *Mind* **111**(444), 751–815. (Cited on pages 201 and 204.)

Chabris, C. F. and D. J. Simons: 1999, 'Gorillas in our midst: sustained inattentional blindness for dynamic events.'. *Perception* **28**(9), 1059–1074. (Cited on page 189.)

Chalmers, D.: 2010, *The Character of Consciousness*. Oxford University Press. (Cited on pages 64, 77, 79, and 81.)

Chalmers, D. and F. Jackson: 2001, 'Conceptual Analysis and Reductive Explanation'. *Philosophical Review* **110**(2), 315–361. (Cited on pages 60, 61, 63, 64, and 185.)

Chalmers, D. J.: 1995, 'Absent qualia, fading qualia, dancing qualia'. In: T. Metzinger (ed.): *Conscious Experience*. Ferdinand Schoningh. (Cited on pages 39, 40, and 41.)

Chalmers, D. J.: 1996, *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, USA, 1 edition. (Cited on pages 3, 4, 5, 6, 7, 8, 10, 12, 39, 42, 45, 49, 56, 60, and 156.)

Chalmers, D. J.: 1997, 'Availability: The cognitive basis of experience?'. In: N. Block, O. J. Flanagan, and G. Guzeldere (eds.): *The Nature of Consciousness*, Vol. 20 ,. Mit Press, pp. 148–149 ,. (Cited on page 12.)

Chalmers, D. J.: 2002, 'Does conceivability entail possibility?'. In: T. S. Gendler and J. Hawthorne (eds.): *Conceivability and Possibility*. Oxford University Press, pp. 145–200. (Cited on pages 45, 46, 47, 48, and 49.)

Chalmers, D. J.: 2003a, 'Consciousness and its place in nature'. In: S. P. Stich and T. A. Warfield (eds.): *Blackwell Guide to the Philosophy of Mind*. Blackwell. (Cited on pages 48 and 103.)

Chalmers, D. J.: 2003b, 'The content and epistemology of phenomenal belief'. In: Q. Smith and A. Jokic (eds.): *Consciousness: New Philosophical Perspectives*. Oxford University Press. (Cited on pages 66 and 80.)

Chalmers, D. J.: 2004, 'The representational character of experience'. In: B. Leiter (ed.): *The Future for Philosophy*. Oxford University Press, pp. 153–181. (Cited on pages 116, 117, 134, and 143.)

Chalmers, D. J.: 2007, 'Phenomenal concepts and the explanatory gap'. In: T. Alter and S. Walter (eds.): *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press. (Cited on pages 68, 73, 75, 77, 78, and 80.)

Chalmers, D. J.: 2009, 'The Two-Dimensional Argument Against Materialism'. In: B. P. McLaughlin and S. Walter (eds.): *Oxford Handbook to the Philosophy of Mind*. Oxford University Press. (Cited on pages 49, 51, and 60.)

Chuard, P.: 2010, 'Non-transitive looks & fallibilism'. *Philosophical Studies* **149**(2). (Cited on pages 87 and 88.)

Churchland, P.: 2005, 'A neurophilosophical slant on consciousness research'. *Progress in Brain Research* **149**, 285–293. (Cited on page 220.)

Churchland, P. M.: 1989, 'Knowing qualia: A reply to Jackson'. In: *A Neurocomputational Perspective*. MIT Press. (Cited on page 55.)

Churchland, P. M.: 1996, 'The rediscovery of light'. *Journal of Philosophy* **93**(5), 211–28. (Cited on page 48.)

Cohen, M. X., S. van Gaal, K. R. Ridderinkhof, and V. A. F. Lamme: 2009, 'Unconscious errors enhance prefrontal-occipital oscillatory synchrony'. *Frontiers in Human Neuroscience* **3**, 54. PMID: 19956401. (Cited on page 216.)

Crane, T.: 1992, 'The nonconceptual content of experience'. In: T. Crane (ed.): *The Contents of Experience*. Cambridge: Cambridge University Press. (Cited on page 181.)

Crick, F. and C. Koch: 1990, 'Towards a neurobiological theory of consciousness'. *Seminars in the Neurosciences* **2**, 263–275. (Cited on page 67.)

Crick, F. and G. Mitchison: 1983, 'The function of dream sleep'. *Nature* **304**, 111–114. (Cited on page 197.)

Crook, S. and C. Gillet: 2001, 'Why physics alone cannot define the 'physical': Materialism, metaphysics, and the formulation of physicalism'. *Canadian Journal of Philosophy* **31**(3), 333–360. (Cited on page 8.)

Cummins, R. E.: 1975, 'Functional analysis'. *Journal of Philosophy* **72**(November), 741–64. (Cited on pages 159 and 160.)

Damasio, A.: 2000, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Mariner Books, 1 edition. (Cited on pages 120, 175, 219, 220, and 225.)

Damasio, A.: 2010, *Self Comes to Mind: Constructing the Conscious Brain*. Pantheon, 1 edition. (Cited on pages 175, 219, 220, 225, and 226.)

Damasio, A. R.: 1995, *Descartes' Error: Emotion, Reason, and the Human Brain*. Harper Perennial, 1 edition. (Cited on page 120.)

Davidson, D.: 1970, 'Mental events'. In: L. Foster and J. W. Swanson (eds.): *Experience and Theory*. Humanities Press. (Cited on page 56.)

Davidson, D.: 1984, *Inquiries Into Truth And Interpretation*. Oxford University Press. (Cited on page 75.)

Davidson, D.: 1987, 'Knowing one's own mind'. *Proceedings and Addresses of the American Philosophical Association* **60**(3), 441–458. (Cited on page 150.)

de Clercq, R. and L. Horsten: 2004, 'Perceptual indiscriminability: In defence of Wright's proof'. *Philosophical Quarterly* **54**(216 ,), 439–444. (Cited on page 87.)

Dehaene, S.: 2009, 'Neural Global Workspace'. In: A. C. Tim Bayne and P. Wilken (eds.): *The Oxford Companion to Consciousness*. Oxford university press. (Cited on page 186.)

Dennett, D.: 1976, 'Are dreams experiences?'. *Philosophical Review* **73**, 151–171. (Cited on page 198.)

Dennett, D. C.: 1991, *Consciousness Explained*. Back Bay Books, 1 edition. (Cited on pages 48 and 185.)

Deutsch, M.: 2005, 'Intentionalism and intransitivity'. *Synthese* **144**(1), 1–22. (Cited on pages 88 and 100.)

Diaz-Leon, E.: 2010a, 'Can phenomenal concepts explain the epistemic gap?'. *Mind*. (Cited on pages 77 and 78.)

Diaz-Leon, E.: 2010b, 'Reductive explanation, concepts, and a priori entailment'. *Philosophical Studies*. (Cited on pages 61 and 64.)

Dretske, F.: 1981, *Knowledge and the Flow of Information*. Cambridge: MIT Press. (Cited on page 181.)

Dretske, F.: 1988, *Explaining Behavior: Reasons in a World of Causes*. MIT Press. (Cited on pages 75 and 146.)

Dretske, F.: 1993, 'Conscious experience'. *Mind* **102**(406), 263–283. (Cited on page 23.)

Dretske, F.: 1995, *Naturalizing the Mind*. MIT Press. (Cited on pages 132, 151, and 210.)

Dummett, M.: 1975, 'Wang's paradox'. *Synthese* **30**(3-4 ,), 201–32. (Cited on pages 84 and 90.)

Egan, A.: 2006a, 'Appearance properties?'. *Noûs* **40**(3), 495–521. (Cited on pages 135, 137, 139, 141, and 166.)

Egan, A.: 2006b, 'Secondary qualities and self-location'. *Philosophy and Phenomenological Research* **72**(1), 97–119. (Cited on page 139.)

Evans, G.: 1982, *The Varieties of Reference*. Oxford University Press, USA. (Cited on pages 115 and 181.)

Fara, D. G.: 2001, 'Phenomenal continua and the sorites'. *Mind* **110**(440), 905–935. (Cited on page 87.)

Feigl, H.: 1958, 'The 'mental' and the 'physical''. *Minnesota Studies in the Philosophy of Science* **2**, 370–497. (Cited on pages 8 and 49.)

Fine, K.: 1975, 'Vagueness, truth and logic'. *Synthese* **54**, 235–259. (Cited on page 84.)

Flanagan, O.: 1995, 'Deconstructing dreams: The spandrels of sleep'. *The Journal of Philosophy* **92**(1), 5–27. (Cited on page 197.)

Flanagan, O. J.: 1993, *Consciousness Reconsidered*. The MIT Press. (Cited on page 175.)

Fodor, J. A.: 1990, *A Theory of Content and Other Essays*. MIT Press. (Cited on pages 75 and 210.)

Foulkes, D.: 1985, *Dreaming: A cognitive-psychological analysis.* Erlbaum. (Cited on page 197.)

Franklin, M. and M. Zyphur: 2005, 'The Role of Dreams in the Evolution of the Human Mind'. *Evolutionary Psychology* **3**, 59–78. (Cited on page 197.)

Fuller, S. and M. Carrasco: 2006, 'Exogenous attention and color perception: performance and appearance of saturation and hue'. *Vision Research* **46**(23), 4032–4047. PMID: 16979690. (Cited on pages 125 and 126.)

Gallagher, S. and D. Zahavi: 2006, 'Phenomenological Approaches to Self-Consciousness'. http://plato.stanford.edu/entries/self-consciousness-phenomenological/. (Cited on pages 21 and 175.)

Gennaro, R. J.: 1996, *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*. John Benjamins. (Cited on pages 201, 203, and 209.)

Godfrey-Smith, P.: 1994, 'A modern history theory of functions'. *NoÃ»s* **28**(3), 344–362. (Cited on pages 159 and 162.)

Goldman, A. and K. Shanton: 2010, 'The Case for Simulation'. In: A. Leslie and T. German (eds.): *Handbook of Theory of Mind*. Psychology Press. (Cited on page 207.)

Goldman, A. I.: 2006, *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press, USA, illustrated edition edition. (Cited on page 207.)

Goodman, N.: 1951, *The Structure of Appearance*. Harvard University Press. (Cited on pages 86 and 90.)

Grahek, N.: 2001, *Feeling pain and being in pain*. Bradford Books. (Cited on page 120.)

Halsey, M. and A. Chapanis: 1951, 'Number of Absolutely Identifiable Hues'. *Jour. Optical Soc. Amer.* **41**(12), 1057–1058. (Cited on page 181.)

Hardin, C. L.: 1997, 'Reinverting the spectrum'. In: A. Byrne and D. R. Hilbert (eds.): *Readings on Color, Volume 1: The Philosophy of Color*, Vol. 1. Mit Press. (Cited on pages 38 and 121.)

Harman, G.: 1990, 'The intrinsic quality of experience'. *Philosophical Perspectives* **4**, 31–52. (Cited on pages 48 and 117.)

Harman, G.: 1996, 'Qualia and Color Concepts'. *Philosophical Issues* **7**, 75–79. (Cited on page 118.)

Hempel, C.: 1969, 'Reduction: Ontological and Linguistic Facets'. In: P. S. Sidney Morgenbesser and M. White (eds.): *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*. St Martin's Press. (Cited on page 8.)

Hill, C. S. and B. P. Mclaughlin: 1999, 'There are fewer things in reality than are dreamt of in Chalmers's philosophy'. *Philosophy and Phenomenological Research* **59**(2), 445–454. (Cited on pages 66, 68, 77, and 80.)

Hinton, J.: 1973, *Experiences*. Oxford: Clarendon Press. (Cited on page 113.)

Hobson, J., E. Pace Schott, and R. Stickgold: 1994, *The chemistry of conscious states*. Little, Brown. (Cited on page 197.)

Hobson, J., E. Pace-Schott, and R. Stickgold: 2000, 'Toward a cognitive neuroscience of conscious states'. *Behavioral and Brain Science* **23**, 793–842. (Cited on page 194.)

Hobson, J. A. and McCarley: 1977, 'The brain as a dream state generator: An activation-synthesis hypothesis of the dream process.'. *American Journal of Psychiatry* **134**, 1335–1348. (Cited on page 197.)

Holman, E. L.: 2002, 'Color eliminativism and color experience'. *Pacific Philosophical Quareterly* **83**(1), 38–56. (Cited on page 134.)

Hume, D.: 1739, *A Treatise of Human Nature*. Oxford University Press, USA. (Cited on page 175.)

Jackson, F.: 1982, 'Epiphenomenal qualia'. *Philosophical Quarterly* **32**(April), 127–136. (Cited on page 53.)

Jackson, F.: 1994, 'Armchair metaphysics'. In: J. O'Leary-Hawthorne and M. Michael (eds.): *Philosophy in Mind*. Kluwer. (Cited on page 7.)

Kalderon, M. and D. Hilbert: 2000, 'Color and the inverted spectrum'. In: S. Davis (ed.): *Color Perception: Philosophical, Psychological, Artistic, and Computational Perspectives*. Oxford University Press. (Cited on pages 38 and 121.)

Kastner, S.: 2010, 'Attention, neural basis'. In: A. C. Tim Bayne and P. Wilken (eds.): *The Oxford companion to consciousness*. Oxford University Press. (Cited on page 229.)

Keefe, R.: 2000, *Theories of Vagueness*. Cambridge University Press. (Cited on page 84.)

Kim, J. and E. Sosa: 1999, *Metaphysics: An Anthology*. Blackwell Publishers. (Cited on page 93.)

Kirk, R.: 1996, *Raw Feeling: A Philosophical Account of the Essence of Consciousness*. Oxford University Press, USA. (Cited on page 23.)

Kraft, J. M. and J. S. Werner: 1994, 'Spectral efficiency across the life span: flicker photometry and brightness matching'. *Journal of american optical society* **11**(4), 1113–1120. (Cited on page 127.)

Kriegel, U.: 2003, 'Consciousness, higher-order content, and the individuation of vehicles'. *Synthese* **134**(3), 477–504. (Cited on page 214.)

Kriegel, U.: 2005, 'Naturalizing Subjective Character'. *Philosophy and Phenomenological Research* **71**(1), 23–57. (Cited on pages 16, 22, and 200.)

Kriegel, U.: 2006, 'Consciousness: Phenomenal Consciousness, Access Consciousness, and Scientific Practice'. In: P. Thagard (ed.): *Handbook of Philosophy of Psychology and Cognitive Science*. Amsterdam: North-Holland. (Cited on pages 14 and 22.)

Kriegel, U.: 2009, *Subjective Consciousness: A Self-Representational Theory*. Oxford University Press, USA. (Cited on pages 9, 15, 16, 22, 23, 183, 184, 201, 204, 209, 210, 213, 215, 216, and 228.)

Kriegel, U.: forthcoming, 'Self-Representationalism and the Explanatory Gap'. In: J. Liu and J. Perry (eds.): *Consciousness and the Self: New Essays*. CUP. (Cited on page 213.)

Kripke, S. A.: 1980, *Naming and Necessity*. Harvard University Press. (Cited on pages 32 and 75.)

LaBerge, S.: 1988, 'Lucid dreaming in Western literature'. In: *Conscious Mind, Sleeping Brain. Perspectives on Lucid Dreaming*. Plenum. (Cited on page 199.)

LaBerge, S., D. W. C. P. Nagel, L. E., and V. P. Zarcone: 1981, 'Lucid dreaming verified by volitional communication during REM sleep'. *Perceptual and Motor Skills* **52**, 727–723. (Cited on page 198.)

Landman, R., H. Spekreijse, and V. A. F. Lamme: 2003, 'Large capacity storage of integrated objects before change blindness'. *Vision Research* **43**(2), 149–164. PMID: 12536137. (Cited on pages 189 and 190.)

Lau, H.: 2008, 'A Higher-Order Bayesian Decision Theory of Perceptual Consciousness'. *Progress in Brain Research* **168**. (Cited on pages 194 and 196.)

Lau, H. and R. Passingham: 2006, 'Relative Blindsight in Normal Observers and the Neural Correlate of Visual Consciousness'. *Proceedings of the National Academy of Science*. (Cited on pages 192 and 193.)

Laureys, S.: 2005, 'The neural correlate of (un)awareness: lessons from the vegetative state'. *Trends in Cognitive Sciences* **9**(12), 556–559. PMID: 16271507. (Cited on page 220.)

Laureys, S. and G. Tononi: 2008, *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology*. Academic Press, 1 edition. (Cited on pages 99 and 225.)

Levine, J.: 1983, 'Materialism and qualia: The explanatory gap'. *Pacific Philosophical Quarterly* **64**(October), 354–61. (Cited on pages 54 and 60.)

Levine, J.: 2001, *Purple Haze: The Puzzle of Consciousness*. Oxford University Press. (Cited on pages 16 and 17.)

Lewis, D.: 1972, 'Psychophysical and Theoretical Identifications'. *Australasian Journal of Philosophy* **50**, 249–58. (Cited on page 37.)

Lewis, D.: 1978, 'Mad pain and Martian pain'. In: N. Block (ed.): *Readings in the Philosophy of Psychology*, Vol. 1. Harvard university Press. (Cited on page 34.)

Lewis, D.: 1979, 'Attitudes de dicto and de se'. *Philosophical Review* **88**(4), 513–543. (Cited on page 138.)

Lewis, D.: 1983, 'Extrinsic properties'. *Philosophical Studies* **44**(2), 197–200. (Cited on page 110.)

Lewis, D.: 1986, *On the Plurality of Worlds*. Blackwell Publishers, first edition edition. (Cited on pages 5 and 6.)

Lewis, D.: 1994, 'Reduction of mind'. In: S. Guttenplan (ed.): *Companion to the Philosophy of Mind*. Blackwell. (Cited on pages 8 and 48.)

Llinas, R. and D. Pare: 1991, 'Of dreaming and wakefulness'. *Neuroscience* **44**, 521–535. (Cited on page 198.)

Llinas, R. R.: 2002, *I of the Vortex: From Neurons to Self*. The MIT Press. (Cited on page 220.)

Loar, B.: 1990, 'Phenomenal states'. *Philosophical Perspectives* **4**, 81–108. (Cited on pages 66 and 119.)

Locke, J.: 1690/1994, *An Essay Concerning Human Understanding*. Prometheus Books. (Cited on page 38.)

Lutze, M., N. J. Cox, V. C. Smith, and J. Pokorny: 1990, 'Genetic studies of variation in Rayleigh and photometric matches in normal trichromats'. *Vision Research* **30**(1), 149–162. PMID: 2321360. (Cited on page 127.)

Lycan, W. G.: 1996, *Consciousness and Experience*. The MIT Press. (Cited on pages 23, 132, 201, and 203.)

Lynn, C., 'Deferential phenomenal concepts? Not for the Zombie Mary.'. (Cited on pages 71 and 72.)

Macpherson, F.: 2006, 'Ambiguous figures and the content of experience'. *Noûs* **40**(1), 82–117. (Cited on page 124.)

Malcolm, N.: 1959, *Dreaming*. Routledge and Kegan Paul. (Cited on pages 194 and 197.)

Maquet, P., J. Peters, J. Aerts, G. Delfiore, C. Degueldre, A. Luxen, and G. Franck: 1996, 'Functional neuroanatomy of human rapid-eye-movement sleep and dreaming'. *Nature* **383**, 163–166. (Cited on page 195.)

Martin, M. G.: 2004, 'The Limits of Self-Awareness'. *Philosophical Studies* **120**, 37–89. (Cited on pages 113 and 114.)

Martinez, M.: 2010, 'A naturalistic account of content and an application to modal epistemology'. Ph.D. thesis, University of Barcelona. (Cited on pages 75, 146, and 149.)

Martinez, M.: 2011, 'Imperative Content and the Painfulness of Pain'. *Phenomenology and the Cognitive Sciences* **10**(1), 67–90. (Cited on page 120.)

Maund, J. B.: 1995, *Colours: Their Nature and Representation*. Cambridge University Press. (Cited on page 134.)

McDowell, J.: 1996, *Mind and World*. Harvard University Press. (Cited on page 112.)

McGinn, C.: 1989, 'Can we solve the mind-body problem?'. *Mind* **98**(July), 349–66. (Cited on pages 5 and 56.)

McLaughlin, P.: 2001, *What Functions Explain. Functional Explanation and Selfreproducing Systems*. Cambridge: Cambridge University Press. (Cited on page 162.)

Merikle, P. M. and M. Daneman: 1999, 'Conscious vs. unconscious perception'. In: M. S. Gazzaniga (ed.): *The new cognitive neuroscience*. The MIT press. (Cited on pages 179 and 180.)

Merikle, P. M., S. Jordens, and J. Stolz: 1995, 'Measuring the magnitude of unconscious influences'. *Consciousness and cognition* **4**, 422–439. (Cited on page 179.)

Metzinger, T.: 2003, *Being No One: The Self-Model Theory of Subjectivity*. The MIT Press, illustrated edition edition. (Cited on page 205.)

Millikan, R.: 1996, 'On Swampkinds'. *Mind and Language* **11**(1), 70–130. (Cited on page 151.)

Millikan, R.: 2000, *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge: Cambridge University Press. (Cited on page 154.)

Millikan, R. G.: 1984, *Language, Thought and Other Biological Categories*. MIT Press. (Cited on pages 146, 149, 159, 205, and 210.)

Millikan, R. G.: 1989, 'Biosemantics'. *Journal of Philosophy* **86**(July ,), 281–97. (Cited on pages 75, 121, 149, and 159.)

Millikan, R. G.: 2002, 'Biofunctions: Two paradigms'. In: A. Ariew (ed.): *Functions*. Oxford University Press. (Cited on pages 161 and 212.)

Montero, B. and D. Papineau: 2005, 'A defense of the via negativa argument for physicalism'. *Analysis* **65**(3), 233–237. (Cited on page 8.)

Mossio, M., C. Saborido, and A. Moreno: 2009, 'An organizational account of biological functions'. *British Journal for the Philosophy of Science* **60**(4), 813–841. (Cited on pages 161, 162, and 163.)

Nagasawa, Y.: 2002, 'The knowledge argument against dualism'. *Theoria* **68**(3), 205–223. (Cited on pages 55 and 57.)

Nagel, E.: 1961, *The Structure of Science: Problems in the Logic of Scientific Explanation*. Harcourt, Brace & World. (Cited on page 160.)

Nagel, T.: 1974/2002, 'What is it like to be a bat?'. In: D. Chalmers (ed.): *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press. (Cited on pages 5, 10, 20, and 21.)

Neander, K.: 1991, 'Functions as selected effects: The conceptual analyst's defense'. *Philosophy of Science* **58**(2 ,), 168–184. (Cited on pages 121, 146, and 149.)

Neander, K.: 1998, 'The division of phenomenal labor: A problem for representationalist theories of consciousness'. *Philosophical Perspectives* **12**(S12), 411–34. (Cited on page 209.)

Neitz, M. and J. Neitz: 1998, 'Molecular Genetics and the Biological Basis of Color Vision'. In: R. K. W.G. Backhaus and J. Werner (eds.): *Color Vision: Perspectives from Different Disciplines*. De Greuter: Berlin. (Cited on page 127.)

Nichols, S. and S. P. Stich: 2003, *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford University Press, USA, illustrated edition edition. (Cited on page 207.)

Nickel, B.: 2006, 'Against intentionalism'. *Philosophical Studies*. (Cited on page 124.)

Pagano, C. and M. T. Turvey: 1998, 'Eigenvectors of the inertia tensor and perceiving the orientations of limbs and objects'. *Journal of Applied Biomechanics* **14**, 331–359. (Cited on page 180.)

Papineau, D.: 1993, *Philosophical Naturalism*. Blackwell. (Cited on pages 75, 95, 121, 146, and 149.)

Papineau, D.: 2002, *Thinking About Consciousness*. Oxford University Press. (Cited on pages 61, 66, 80, 95, 96, and 97.)

Peacocke, C.: 1984, *Sense and Content: Experience, Thought, and Their Relations*. Oxford University Press, USA. (Cited on pages 119 and 123.)

Peacocke, C.: 1986, 'Analogue content'. *Proceedings of the Aristotelian Society* **60**, 1–17. (Cited on page 181.)

Peacocke, C.: 1995, *A Study of Concepts*. The MIT Press. (Cited on page 205.)

Perry, J.: 2001, *Knowledge, Possibility and Consciousness*. Cambridge, MIT Press. (Cited on page 66.)

Pietroski, P. M.: 1992, 'Intentionality and teleological error'. *Pacific Philosophical Quarterly* **73**(3 ,), 267–82. (Cited on page 154.)

Pineda, D.: 2006, 'A mereological characterization of physicalism'. *International Studies in the Philosophy of Science* **20**(3), 243–266. (Cited on page 8.)

Pollen, D. A.: 2008, 'Fundamental Requirements for Primary Visual Perception'. *Cerebral Cortex* **18**(9), 1991–1998. (Cited on page 175.)

Prinz, J.: Forthcoming, 'Is Attention Necessary and Sufficient for Consciousness?'. In: D. S. Christopher Mole and W. Wu (eds.): *Attention: Philosophical and Psychological Essays*. Oxford University Press. (Cited on page 229.)

Prinz, J. J.: 2004, *Gut Reactions: A Perceptual Theory of Emotion*. Oxford University Press, USA, illustrated edition edition. (Cited on page 120.)

Putnam, H.: 1975, 'The meaning of 'meaning''. *Minnesota Studies in the Philosophy of Science* **7**, 131–193. (Cited on pages 69, 76, and 131.)

Quine, W. V. O.: 1948, 'On what there is'. *Review of Metaphysics* **2**, 21–38. (Cited on page 43.)

Revonsuo, A.: 1995, 'Consciousness, Dreams, and Virtual Realities'. *Philosophical Psychology* **8**, 35–58. (Cited on page 198.)

Revonsuo, A.: 2000, 'The Reinterpretation of Dreams: An evolutionary hypothesis of the function of dreaming'. *Behavioral and Brain Sciences* **23**(6), 877–901. (Cited on pages 194 and 197.)

Revonsuo, A. and J. B. Newman: 1999, 'Binding and consciousness'. *Consciousness and Cognition* **8**(2), 123–127. (Cited on page 4.)

Rey, G.: 1995, 'Toward a projectivist account of conscious experience'. In: T. Metzinger (ed.): *Conscious Experience*. Ferdinand Schoningh. (Cited on page 48.)

Roffwarg, H. P., J. N. W.C. Dement, J.N. Muzio, and C. Fisher: 1962, 'Dream imagery: Relationship to rapid eye movements of sleep'. *Archives of General Psychiatry* **7**, 235–258. (Cited on page 198.)

Rosenthal, D.: 2008, 'Consciousness and its function'. *Neuropsychologia* **46**(3), 829–840. (Cited on page 194.)

Rosenthal, D. M.: 1986, 'Two concepts of consciousness'. *Philosophical Studies* **49**(May), 329–59. (Cited on pages 10 and 11.)

Rosenthal, D. M.: 1997, 'A theory of consciousness'. In: N. Block, O. J. Flanagan, and G. Guzeldere (eds.): *The Nature of Consciousness*. Mit Press. (Cited on pages 11, 23, 184, 187, 201, and 203.)

Rosenthal, D. M.: 2005, *Consciousness and mind*. Oxford University Press. (Cited on pages 20, 22, 23, 177, 184, 187, 201, 203, and 209.)

Schnall, S., Haidt, and G. J., Clore: 2008, 'Disgust as embodied moral judgment.'. *Personality and Social Psychology Bulletin* **34**(8), 1096–1109. (Cited on page 180.)

Schroeder, T.: 2004, 'New norms for teleosemantics'. In: H. Clapin (ed.): *Representation in Mind*. Elsevier. (Cited on page 161.)

Searle, J. R.: 1992, *The Rediscovery of the Mind*. MIT Press. (Cited on page 75.)

Shoemaker, S.: 1982, 'The inverted spectrum'. *Journal of Philosophy* **79**(July), 357–381. (Cited on page 38.)

Shoemaker, S.: 1990, 'Qualities and qualia: What's in the mind?'. *Philosophy and Phenomenological Research Supplement* **50**(Supplement), 109–131. (Cited on page 134.)

Shoemaker, S.: 1994, 'Phenomenal character'. *Noûs* **28**(1), 21–38. (Cited on pages 133 and 134.)

Shoemaker, S.: 2000, 'Phenomenal character revisited'. *Philosophy and Phenomenological Research* **60**(2), 465–467. (Cited on page 136.)

Shoemaker, S.: 2001, 'Introspection and Phenomenal Character'. *Philosophical Topics* **28**(2), 247–273. (Cited on page 117.)

Smart, J. J. C.: 1978, 'The content of physicalism'. *Philosophical Quarterly* **28**(October), 339–41. (Cited on page 8.)

Smith, A. D.: 2002, *The problem of perception*. Harvard University Press. (Cited on page 113.)

Solms, M.: 1997, *The neuropsychology of dreams: A clinico-anatomical study*. Erlbaum. (Cited on page 197.)

Sorensen, R. A.: 2001, *Vagueness and Contradiction*. Oxford University Press. (Cited on pages 85 and 158.)

Sperling, G.: 1960, 'The information available in brief visual presentation'. *Psychological Monographs: General and Applied* **74**(11), 1–29. (Cited on page 188.)

Stalnaker, R. C.: 1999, *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford University Press, USA. (Cited on page 138.)

Stoljar, D.: 2005, 'Physicalism and phenomenal concepts'. *Mind and Language* **20**(2), 296–302. (Cited on pages 49, 65, and 69.)

Tononi, G.: 2009, 'Sleep and Dreaming'. In: S. Laurey and G. Tononi (eds.): *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology*. Elsevier. (Cited on pages 195 and 199.)

Tye, M.: 1990, 'Vague objects'. *Mind* **99**(396), 535–557. (Cited on page 84.)

Tye, M.: 1996, 'Is consciousness vague or arbitrary?'. *Philosophy and Phenomenological Research* **56**(3), 679–685. (Cited on pages 20, 97, 99, and 157.)

Tye, M.: 1997, *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. The MIT Press. (Cited on pages 23, 25, 87, 120, 124, 132, and 184.)

Tye, M.: 1999, 'Phenomenal consciousness: The explanatory gap as a cognitive illusion'. *Mind* **108**(432), 705–25. (Cited on pages 59, 68, and 69.)

Tye, M.: 2002, *Consciousness, Color, and Content*. The MIT Press. (Cited on pages 23, 25, 87, 129, 152, and 184.)

Tye, M.: 2003a, 'Blurry images, double vision, and other oddities: New problems for representationalism?'. In: Q. Smith and A. Jokic (eds.): *Consciousness: New Philosophical Essays*. Oxford University Press. (Cited on page 124.)

Tye, M.: 2003b, 'A theory of phenomenal concepts'. In: A. O'Hear (ed.): *Minds and Persons*. Cambridge University Press. (Cited on pages 66 and 69.)

Tye, M.: 2009, *Consciousness Revisited: Materialism without Phenomenal Concepts*. The MIT Press. (Cited on pages 68, 69, and 71.)

Van Gulick, R.: 2004, 'Higher-order global states (hogs): An alternative higher-order model of consciousness'. In: R. J. Gennaro (ed.): *Higher-Order Theories of Consciousness: An Anthology*. John Benjamins. (Cited on pages 201 and 209.)

Weatherson, B.: 2006, 'Intrinsic vs. Extrinsic Properties'. http://plato.stanford.edu/entries/intrinsic-extrinsic/. (Cited on page 110.)

Weiskrantz, L.: 1986, *Blindsight: A Case Study and Implications*. Oxford University Press USA, 1 edition. (Cited on pages 12 and 179.)

Whalen, P., S. Rauch, N. Etcoff, S. McInerney, M. B. Lee, and M. Jenike: 1998, 'Masked presentation of emotional facial expessions modulate amygala activity without explicit knowledge'. *Jounal of Neuroscience* **18**, 411–418. (Cited on page 179.)

Williams, A.: 1938, 'Perception of subliminal visual stimuli'. *Journal of Psychology* **6**, 187–199. (Cited on page 179.)

Williamson, T.: 1996, *Vagueness*. Routledge. (Cited on pages 85 and 158.)

Winson, J.: 1993, 'The biology and function of rapid eye movement sleep'. *Current Opinions in Neurobiology* **3**, 243–248. (Cited on page 196.)

Wittgenstein, L.: 1968, 'Notes for lectures on private experience and sense data'. *Philosophical Review* **77**(July), 275–320. (Cited on page 38.)

Wouters, A.: 2005, 'The function debate in philosophy'. *Acta Biotheoretica* **53**(2), 123–151. (Cited on pages 148 and 159.)

Wright, C.: 1975, 'On the coherence of vague predicates'. *Synthese* **30**(3-4), 325–65. (Cited on page 86.)

Wright, L.: 1976, *Teleological Explanations*. University of California Press. (Cited on page 148.)

Wright, W.: 2003, 'Projectivist representationalism and color'. *Philosophical Psychology* **16**(4), 515–529. (Cited on page 134.)

Yablo, S.: 1993, 'Is Conceivability a Guide to Possibility?'. *Philosophy and Phenomenological Research* **53**(1), 1–42. (Cited on page 46.)

COLOPHON

This thesis was typeset in LyX, available from www.lyx.org. The typographic style is available for LaTeX via CTAN as "`classicthesis`". A classicthesis port for LyX is available from www.soundsorange.net.

Some figures were created using the GIMP Biggles module for Python (gimp.org), others using OpenOffice (www.openoffice.org).

Final Version as of May 2, 2011 at 17:10.