Evolution’s Shorthand
A Presentational Theory of the Phenomenal Mind

een wetenschappelijke proeve op het gebied van de wijsbegeerte

Proefschrift

ter verkrijging van de graad van doctor aan de Katholieke Universiteit Nijmegen, op gezag van de Rector Magnificus Prof. dr. C.W.P.M. Blom, volgens besluit van het College van Decanen in het openbaar te verdedigen op maandag 14 april 2003, des namiddags om 3.30 uur precies door

Hans Dooremalen

ter verkrijging van de graad van doctor aan de Katholieke Universiteit Nijmegen, op gezag van de Rector Magnificus Prof. dr. C.W.P.M. Blom, volgens besluit van het College van Decanen in het openbaar te verdedigen op maandag 14 april 2003, des namiddags om 3.30 uur precies door

Hans Dooremalen

geboren op 8 september 1967
te Rijen.
Promotor:
Prof. dr. A.A. Derksen

Manuscriptcommissie:
Prof. dr. J.A.M. Bransen
Prof. dr. J. Leilich (Universiteit Antwerpen – UFSIA)
Dr. M.A.M.M. Meijsing (Universiteit van Tilburg)

© 2003 Hans Dooremalen
ACKNOWLEDGMENTS
Writing a dissertation would not be possible without a great deal of help. Many people have contributed in various ways and different stages to this enterprise. I would like to thank them all.

First of all, Ton Derksen provided me with the opportunity to do research in the interesting area of phenomenal consciousness. I have benefited enormously from his extensive comments on the various drafts of all the chapters of this dissertation.

The position I have developed in this dissertation also owes a lot to the discussions I had with Marc Slors and the suggestions he made. I also want to thank him for being the subject of the only empirical experiment I have ever done as a philosopher.

Many of my ideas have been inspired by the – often informal – discussions in the research group Rationality and Non-Reductionism. Besides Ton and Marc, I would like to thank the other participants: Chris Buskens, Anthonie Meijers, Monica Meijsing, Herman de Regt, Esteban Rivas, Pouwel Slurink, and Arno Wouters. It has been a privilege to listen to you.

Special thanks to Filip Buekens. It is simply not conceivable that I could have written this dissertation without his contagious enthusiasm for analytical philosophy having infected my mind.

I owe a special debt to Chiel Bertens en Anique Lensen. Being close to me both emotionally and in space-time they were in turns the first to hear what I had thought up. Thank you for listening and responding to me.

Thanks to Vincent Kwantes and Wannes Rombouts for all the lively discussions about every topic in this dissertation and many, many more. Apologies for all the times I rattled on while we should have been rocking.

Thanks to Frank Scheele for drawing my attention to tetrachromacy.

No idea is worth being developed if it isn’t shared. I like to thank my friends: Boudewijn Aerts, Olga Ansems, Bregtje Bertens, Hedwig Spadaccini-Zeedijk, Piet-Hein Clijsen, Sander Colijn & Laura van Driel, Leon van Egmond, Peter & Sharon van Gestel, Janneke van Gog, Gerard van Gorkum, Sylvie van Halder, Liane den Hartog, Renske & Frank Henzel, Kiki van der Horst,
Renate Hovius, Esther de Kok, Cyrano Opdebeke, Ron Peek, Myriam & Johan van Rooijen-Horsten, Renske Somerwil en Lex van Tilburg, Thank you all for patiently listening to my lectures in bars, at parties or wherever I saw a chance to tell you about your own but mostly other minds.

I also extend my thanks to others who have contributed in various ways: Jan Bransen, Daniel Cohnitz, Joachim Leilich, Jeroen Linssen, Peter Reijnaert, Johan Veldeman and Teresa Viader-Rapp.

Finally, I want to thank my parents, my brother and his wife for always supporting my decisions in life, even though – be it refusing to serve in the army, studying philosophy, or writing this dissertation. I could never have accomplished that without your support and love.
INTRODUCTION

This dissertation is about the scientific study of conscious experiences. If one takes a look at the contributions to recent conferences on consciousness like the Tucson Conferences and those organized by the Association for the Scientific Study of Consciousness one cannot escape the impression that there seems to be a growing consensus among scientists and philosophers that such a science is possible and that exciting scientific discoveries about the phenomenal mind have been made in the last decade. Even though consensus seems to be growing, there are some philosophical views on consciousness that deny that a science of consciousness is possible. I will argue against these views. I will present Owen Flanagan’s natural method, as the way to find answers to questions about consciousness. Applying this method to different kinds of experiences leads to the theory that I call the evolutionary shorthand theory of the phenomenal mind. This theory provides an answer to the question why phenomenal consciousness was selected in evolution; it is a quick way of picking up information that is important to the survival and reproduction of the experiencing subject. My goal in this dissertation is to support the theory by recent empirical data.

In chapter one I will present my two assumptions, viz. qualia realism and broad physicalism, and the two questions that together form the problem of phenomenal consciousness: How is phenomenal consciousness realized and what is it’s evolutionary function?

Chapter two presents the evolutionary shorthand theory of phenomenal consciousness. There I give a characterization of the phenomenal mind. The chapter is an exposition of the theory. The defense of this theory follows in the rest of the dissertation and consists of (1) arguments against different reasons to think that a scientific solution of the problem of the phenomenal mind is impossible; (2) a presentation and discussion of empirical data concerning five varieties of phenomenal experiences; (3) a critique of arguments against a representationalist account of phenomenal consciousness.

In chapters three and four I argue against the most recent view of tough dualism. Under tough dualism every variety is classified that denies that the phenomenal belongs to the broadly physical domain. If tough dualism is right, a scientific theory of the phenomenal mind is impossible, for science is only concerned with the broadly physical domain. The conceivability argument is the tool most used by the tough dualists. In chapter three I will demonstrate that this kind of argument begs the question if it is applied to something we are ignorant about. In chapter four I argue against the tough dualist position itself. Basically the problem is a variety of the interaction problem; if
the phenomenal mind is not physical and it does have influence on our behavior, how can this be explained without giving up either tough dualism or the claim that the phenomenal has influence on the physical?

Some philosophers are agnostic about accepting either tough dualism or monism. Their argument that we have to remain agnostic, starts with the observation that scientific concepts are objective and that conscious experiences are subjective. The question here is: How can we have an objective science about something that is subjective? It seems that we could never have the right scientific concepts: how can an objective theory ever explain what it is like to be a human? Those who do believe we do not have an answer to this question, but do not exclude that we will find it some day, remain agnostic. In chapter five I will argue that this question poses a demand on a scientific theory of consciousness that is too strong.

Not only tough dualists and agnostic philosophers argue against the possibility of a science of the phenomenal mind, there is also an argument coming from the monist camp: maybe we are not intelligent enough – cognitively closed – to understand phenomenal consciousness and its relations to the rest of the broadly physical domain. This argument is discussed and rejected in chapter six. One of the arguments against cognitive closure comes from Owen Flanagan. He argues that the natural method provides a way to understanding consciousness. I agree. This method is presented in chapter six, and will be applied to various kinds of phenomenal experience in the chapters seven through eleven. I use these case studies to present different aspects of the evolutionary shorthand theory of phenomenal consciousness as presented in chapter two.

Though many aspects recur in the chapters describing the case studies, each of these chapters has one aspect of my characterization of the phenomenal mind that is central.

Chapter seven focuses on projection. Using phantom limbs I will show that phenomenal experiences are projections. In the discussion following the presentation of the data the evolutionary shorthand theory is compared with Max Velmans’ recursive monism that also defends a projectionst view on the phenomenal mind. I argue that contrary to his view phenomenal experiences are located in the head.
The central aspect of chapter eight on synesthesia is that phenomenal experiences are presentations that can be used as representations. Here the main discussion is with Michael Tye. I mainly disagree with his claim that all phenomenal experiences are representations.

In chapter nine the central aspect is the interpretation of presentations as representations. The case study concerns the phenomenon of dreaming. This chapter is also meant as an illustration of the natural method, for much of the data are derived from Owen Flanagan’s book *Dreaming Souls* (2000) in which he successfully uses this method to gain knowledge about dreaming.

The focus of attention of chapter ten is the secondary quality view of phenomenal experiences, i.e. the claim that no object really is colored, sweet, or beautiful. I will defend this claim for colors against the view of Frank Jackson.

In chapter eleven the aspect that is central is the species-specific indirect representation of bodily and environmental features. It is here that I demonstrate that phenomenal presentations can be used to represent features of the environment that are advantageous to the representing organism’s survival and reproduction. Since this way of representing is fast and indirect I call it evolution’s shorthand.

In chapter twelve I discuss three problems that any representationalist theory of the phenomenal mind should face. First, a general threat to representationalism will be rebutted: since phenomenal representations create intensional contexts, phenomenal representations cannot be reduced to nonphenomenal representations that express the same content. The remaining discussion will be with a philosophers which I suspect to be a defender of conscious inessentialism (problem two), Max Velmans, and with those who defend representation inessentialism (problem three; the view that all our actions can be performed without conscious representations) Kevin O’Regan and Alva Noë. The evolutionary shorthand theory of consciousness shows that both kinds of inessentialism are incorrect and comes out as the view which is best in line with the empirical data.
I really like *Chimay Grande Réserve*. Sometimes this causes me to drink too much, which results in a severe hangover the next afternoon. This, I believe, is the shortest argument for the claim that phenomenal consciousness is real and that it causally interacts with the physical world. Together with the scientific commitments to the causal closure of the physical world and the rejection of causal overdetermination, it follows that all phenomenal states are part of the physical world. It also shows that normally we believe that it is the beer that tastes so good and makes us feel good. That is the hedonic reason why I buy and drink it even though I am consciously aware of the negative impact it has on my behavior, perceptual and cognitive capacities and on my future qualia: *I want it and I want it now*. But what is it that I want? Can we give an account of what qualia are, why we have them, and how they are related to the rest of the world?

In this chapter I will do two things. First, I will present the two main assumptions that I accept throughout this dissertation, viz. qualia realism and broad physicalism. Secondly, I will describe the two serious parts of problem of phenomenal consciousness are: why is there phenomenal consciousness? And how is phenomenal consciousness realized?

1. **the two main assumptions: qualia realism and broad physicalism**

In asking the kind of questions I just posed and trying to answer them, I assume the truth of two positions: qualia realism and broad physicalism. Phenomenal consciousness is real: some of our mental states have a what-it-is likeness, a certain qualitative feel. This is not an issue that is open for debate as far as I am concerned. I try to avoid painful situations and pursue pleasurable ones, something creatures without phenomenal experiences can and will not do. For instance, my phenomenal zombie-twin – a hypothetical being that is a physical but not a phenomenal duplicate of me – will have no reason to go to a bar to have a beer with his friends, he will only see things in a blur, will have trouble thinking and walking straight and the next day his body will be sick. Humans with phenomenal experiences sometimes take all this for granted – even the pounding headache the next day – because there are some seriously enjoyable qualia to be found in bars. Fortunately for my zombie-twin, there is nothing it is like for him to have a hangover. It is something it is like to be a drinker, and that’s enough to start with. Here, I rest my defense of qualia-realism.
Being a *qualia-realist* I should give at least a provisional answer to the questions that arise from both parts of this term: *what am I a realist about? And what do I mean by the claim that something is real?* First of all, it should be noted that the qualia-realist is not committed to a *restricted definition* of qualia. On this account qualia are defined as nonrepresentational, not functionally analyzable, or not causally efficacious. I believe it is a bad idea to start off with such a definition, for qualia are real and there might not be anything real corresponding to such a restricted definition. It is enough to say that qualia are the what-it-is-likeness of experiences. Scientific data, not intuition, should provide us with a more elaborate description of what the properties of qualia are. Chapter two starts with a description based on empirical data.

Usually phenomenal consciousness is distinguished from access consciousness (Block 1995) or psychological consciousness (Chalmers 1996b). If this distinction is the distinction between those conscious states that are functionally analyzable and those that are not, I do not believe it is a useful distinction, for then I would have to say that I doubt whether phenomenal consciousness exists: it is not *a priori* evident that phenomenal consciousness cannot be functionally analyzed. The distinction is more useful as it is meant to distinguish between those conscious states that essentially have a what-it-is-likeness and those that have not. Even though I believe that every conscious state has a phenomenal feel (and it might even turn out that this is necessarily so), some of the conscious states might well have no phenomenal feel and still be conscious states. Believing that it rains is such a state that does have a phenomenal feel (which is different from – say – hoping that it rains) but *conceivably* it would still be a conscious state if there were no phenomenal feel to accompany it. This is in contrast with the conscious experience of red, which essentially is phenomenal. Leave this out, and the whole conscious state has been left out.

To the question about what I mean by the claim that phenomenal experiences are *real*, I can give a very short answer: they belong to the *broadly physical domain*. But now the question arises what does it mean that a property or object belongs to the broadly physical domain? By a *broadly* physical property I mean that that property is constituted by physical entities. Biological properties for instance are broadly physical, for they are realized by molecules, which evidently are physical entities. So, if a property belongs to the broadly physical domain, it is either itself physical (in the narrow sense, e.g. the way molecules are said to be physical), or it is constituted by it. This logically entails that all broadly physical properties also have a location. This then is a *monist* view on the world: everything in the world that exists is eventually constituted by physical
'stuff.' The reason for calling this view physicalist lies in the fact that we can have physical properties without biological or other higher-order properties, but not vice versa.

Are there also properties that fall outside the broadly physical domain? In chapter four I will argue that we can only know properties that belong to the broadly physical domain. This means that if there are other properties we do not and cannot know them, hence we cannot affirm that they exist. This leads to agnosticism with respect such properties. Since we do know that we are phenomenally conscious it follows that phenomenal consciousness is a broadly physical property.

Some philosophers like John Searle (1992) classify the broadly physical properties as physical themselves. One might also be agnostic about whether the lowest level still can be called physical, which results in a neutral view on monism: if the lowest level is not physical we do not know what to call it, but the assumption will be that there is just one kind of “stuff” that realizes the physical properties in our world, which in turn realize the mental properties in our world.

This kind of neutral monism should not be confused with the kind of monism that says that there is one kind of “protostuff” which is neither physical nor mental, and that this constitutes the particulars of these otherwise very different ontological classes, which would leave the option open that there is another world where this protostuff realizes only mental but no physical properties or vice versa.

Broadly physical properties can be characterized by saying that they all have causal powers. In chapter four I will argue that we only can know and understand those kinds of existent things that do have causal powers. This means that we should be agnostic about accepting entities in our ontology that lack causal powers. Within a broadly physicalist framework such entities are not acceptable, hence to belong to the domain of the broadly physical is the same as to belong to the realm of the causally efficacious. Since we cannot know any noncausal entities, every noncausal entity that was ever discussed was merely a postulate.

Again, this raises a question: does this imply reductionism, where only the – yet unknown – fundamental properties of the universe are causally efficacious? The answer is no: being a physicalist does not imply being a reductionist: higher-order properties are just as real as fundamental properties. Hence, one of the assumptions I make is that nonreductive physicalism is
correct. With this comes the acceptance of a mereological or layered model of the world. I will discuss the mereological model and nonreductive physicalism in turn.

1.1 the mereological model

I am aware of the fact that talk about layers, levels or orders is largely metaphorical. Maybe we should use terms that describe amounts of complexity and functional relations between properties. I use the mereological terminology as an economic way of doing just that. Having said this, we can go about using the model. The world consists of different layers or levels: e.g. the level of atoms, the level of molecules and the level of liquids.

Jaegwon Kim claims that this view on physicalism also implies the acceptance of mind-body supervenience:

\[\text{Mind-body supervenience can serve as a useful dividing line: it can be viewed as defining minimal physicalism. (Kim 1998: 15)}\]

Jackson holds a similar view. He argues that when we do serious metaphysics we should not try to make a list of everything that exists. Instead we should make a small list of fundamental entities (1998: 4). A physicalist account of our world does just that. Everything that is real but which is not explicitly listed in this account has to be “entailed” by it. (Here I use “entailed” as an ontological notion, as Jackson does.) Though these claims are about a physicalist theory, they evidently are relevant to the ontology, since the theory is about our world. Physicalism claims that a complete account of our world can be given in terms of the properties and relations between them that belong to the small list of fundamental entities. It further claims that the physical nature of our world determines the nature of our world without reminder. This means that “[a]ny world that is a minimal physical duplicate of our world is a duplicate simpliciter of our world.” (Jackson 1994: 104) Let us call this the minimal physical duplicate thesis.

According to Jackson materialism is a claim about the actual world – hence the phrase “our world” (1998: 11) and any physical duplicate has to be a duplicate without adding nonphysical stuff – hence the phrase “minimal physical duplicate” (1998: 12). Jackson (1994: 104-105) argues that every materialist is committed to this claim. Suppose that the duplicate world does have
some nature that our world does not have. Then this would be a nonphysical nature, for both worlds are physically identical. However, the duplicate world cannot contain any nonphysical nature by definition, for we only duplicated the physical properties. On the other hand, our world might contain some nonphysical nature. If this is true, then materialism is evidently false. Hence, a materialist has to defend this claim.

In turn this thesis implies the following:

Let \( \Phi \) be the story as told in purely physical terms, which is true at the actual world and all the minimal physical duplicates of the actual world, and false elsewhere. \( ? \) is a hugely complex, purely physical account of our world. Let \( ? \) be any true sentence which is about the psychological nature of our world in the sense that it can only come false by things being different psychologically from the way they actually are: every world at which \( ? \) is false differs in some psychological way from our world. Intuitively, the idea is that \( ? \) counts as being about the psychological nature of our world because making it false requires supposing a change in the distribution of psychological properties and relations. Now, if the minimal physical duplicate thesis is true, every world at which \( \Phi \) is true is a duplicate simpliciter of our world, and so a fortiori a psychological duplicate of our world. But then every world at which \( \Phi \) is true is a world at which \( ? \) is true – that is, \( \Phi \) entails \( ? \). (1998: 25)

This is what Jackson calls the entry by entailment thesis. Clearly, Jackson uses the notion of “entailment” not as a logical notion, in the sense that \( ? \) can be deduced from \( \Phi \). Here, entailment is used to indicate that statements that are part of \( \Phi \) are made true by the same properties and their relations that are described by \( \Phi \) (1998: 25). Since it is the properties that make the statements true, this is also an ontological claim: the claims referred to by the terms of the psychological theory are made true by the ontological entities in the world that can be described by the physical theory. If the psychological claims are true, then the entities described by this theory have to exist and have to be determined by some of the entities described in the physical theory.

I agree with all of the aforementioned claims: all physicalists have to concede to minimal physicalism or a general supervenience thesis, which implies a mereological view of the world. Nonfundamental properties are supervenient on the fundamental properties and cannot vary independently of these fundamental properties.

1.2 nonreductive physicalism

Yet, being a physicalist does not imply that one is a reductionist. It is because that a group of H₂O molecules interact in manner \( X \) with each other at the level of molecules, that water is a liquid at a higher level. If the same molecules were to interact in manner \( Y \), then the water would be ice. This means that liquidity cannot be reduced to the lower level, for there is no liquidity at
this level, or as John Searle puts it: “one cannot reach into a glass of water and pick out a molecule and say ‘This one is wet’ [.]” (1994: 181)

Another argument against reductionism works the other way round: not only can the same lower level entities constitute different higher level properties preventing reduction of higher-level properties to lower-level properties, but the same higher-level properties can also be realized by different lower-level properties, e.g. liquidity can also be realized by copper, iron or gold molecules, which means that there is not one specific lower-level property to reduce the phenomenon to. One might argue that it is functional properties – of say water and copper molecules – that the phenomenon of liquidity can be reduced to. But this will not work, for even if it is exactly the same set of relations between water and copper molecules that constitute liquidity, water and copper molecules are very different, and the reduction will stop there at the level of molecules, which is not enough for reductionism, where only the lowest level properties count.

As I argued earlier, the kind of physicalism I endorse is also a kind of monism. This, however, does not imply that there are no non-physical properties. Biological properties are not themselves physical, but they are broadly physical. The same goes for social or phenomenal properties: they are not physical themselves but are constituted by physical properties, hence they are broadly physical. The broadly physical domain, thus, is a domain that contains many different kinds of properties that are all constituted in the end by physical properties and entities. Property pluralism and monism, then, go hand in hand in this nonreductive physicalism.

Nonreductive physicalism does not imply that there cannot be reductive explanations. I agree with David Chalmers (1996a; 1996b) that we should make a distinction between reduction and reductive explanation. Chalmers sees reduction as identification (1996b: 43) as does Max Velmans (2000: 3). Most natural phenomena are not reducible, liquidity being one of them as I just argued. According to Chalmers a reductive explanation starts with a functional analysis of the explanandum, after which an explanation has to be given in causal terms. A reductive explanation, then consists in answering two questions: (1) a question about what the phenomenon does: what is the function of the phenomenon to be explained?, and (2) a question about the execution of that function: how is this function realized? Suppose we want to explain the biological phenomenon of being a heart. First, we have to answer the question: what is the function of a heart? The answer lies along the following lines: a heart pumps the blood in an organism in order for
oxygen and other chemicals in the blood to reach their destination in the body of the organism. The second question is: how does the heart do this? A long explanation follows about contraction of muscle-tissue, signals that are being sent via neurons from the brain to the heart, etcetera. The level of the functional analysis (the heart) is one that is higher than the level of explanation (the cells from which the heart is build), hence this is a reductive explanation, where no reduction is involved: the heart is as real as is the muscle tissue that constitutes it.

I claimed that having causal powers is the same as being broadly physical. I will use Victor Johnston’s (1999) example to indicate how higher-level properties can have causal influence (in the course of evolution), without falling into the trap of causal overdetermination. He uses a car analogy (1999: 9 ff). A car has been made in factory and has to complete a course that leads back to the factory. Only if the car is able to make the sharp turns and can accelerate up a hill, the car can get back to the factory. Hence, the only cars that make it back home are those who have the higher-order properties of acceleration and the cornering ability, regardless of how these are realized by the parts that constitute the car. So even though these parts constitute the car and do most of the work, it is the higher-order properties that emerge from or supervene on the arrangement of the parts that influence whether the car gets back to the finish line.

Johnston extends this example to evolution. Suppose designers only use (combinations of) the designs again from the cars that made it back to the factory to create new cars. In that case the higher-order properties of acceleration and cornering ability have causal influence on which lower-level properties get passed on to next generations. And these lower-level properties in turn constitute the bases for the emergence of the future higher-order properties of the next generation of cars. The same goes for the relation between consciousness and the brain: “Over the long course of evolution, the functional attributes of the mind have been responsible for shaping the physical and chemical structure of brains. From this viewpoint, functional consequences dictate structural design.” (Johnston 1999: 11) Johnston presents this evolutionary functionalism (1999: 9) as part of the solution to the problem of consciousness (1999: 11). I agree that it is, but what is this problem?

2. the problem of phenomenal consciousness

I believe that the problem of phenomenal consciousness is the a combination of two questions: (1) why do we have phenomenal consciousness given our physical make-up? And (2) why did phenomenal consciousness evolve?
The first question might be taken as expressing a problem that stems from a combination of the dualist intuition most people have and conscious inessentialism, where

Conscious inessentialism is the view that for any intelligent activity \( i \), performed in any cognitive domain \( d \), even if we do \( i \) with conscious accompaniments, \( i \) can in principle be done without these conscious accompaniments. (Flanagan 1992: 5)

Conscious inessentialism is based on the physicalist view that everything in the world is done by (broadly) physical entities and properties. If this is combined with consciousness realism and the view that consciousness is not physical nor broadly physical, a tough kind of dualism results: it is just a small step from the claim that consciousness exists but is not (broadly) physical to the claim of full-blown substance dualism. One only needs to add that it does not depend on the (broadly) physical in any way. (There are several reasons why one would accept conscious inessentialism. In this dissertation some of them are discussed and refuted. Cf. chapter three, six, and twelve.)

If such a position would be correct, then the problem arises: If all the work is done by the (broadly) physical properties in the world, then why is some (or maybe all) of this work accompanied by phenomenal experiences? Nowadays this problem is mostly referred to as the ‘hard problem,’ a phrase made famous by David Chalmers (1993; 1995; 1996a; 1996b; 2002). As I will argue in chapter four, if one puts the question like this, then one already accepts epiphenomenalism, which renders the question insolvable. So, the first question is not a formulation of the hard problem of consciousness.

Contrary to Chalmers’ view, the first question is a question that has to be interpreted within the context of broad physicalism: how is the broadly physical property of phenomenal consciousness realized by the physical? This is a genuine part of the problem of phenomenal consciousness: if phenomenal consciousness is a real property it has to be a property that belongs to the broadly physical world. The question then becomes the one Colin McGinn famously formulated: “How can technicolour phenomenology arise from soggy gray matter?” (1991: 1) This interpretation of the first question clearly differs from the hard problem, for in this version of the problem it is accepted that the phenomenal mind belongs to the broadly physical domain, something which is denied in the interpretation according to Chalmers.

To this question I will present only a partial answer, for I believe that neurology plays a central role here, and at this moment the neurology of phenomenal experiences is far from complete. I
will discuss some neurological findings that are relevant to this question and that hint at how some kinds of phenomenal experiences are realized in the brain.

The second question is an evolutionary one. *Why did phenomenal consciousness evolve? What is its evolutionary function?* I hope that my answer to this question will be satisfying for those who are both realists about phenomenal consciousness and are convinced that it has causal efficacy (I will argue for this latter claim in chapter four, by refuting Chalmers’ position).

In taking an evolutionary look at phenomenal consciousness it immediately becomes clear that we can talk about phenomenal consciousness in general, though there are many different phenomenal conscious states. In other words, phenomenal consciousness is both *homogeneous* and *heterogeneous* at the same time: even though experiences of red, pain or beauty are very different experiences, they all have a what-it-is-likeness. This heterogeneity makes answering questions about phenomenal consciousness hard, for two problems surface immediately. The first problem is how to distinguish the experiences that were selected for from those that were not. If phenomenal consciousness is heterogeneous, then probably not every kind of phenomenal experience was selected for. Experiencing a pleasant taste while eating a ripe strawberry is a candidate for selection, but experiencing beauty while listening to Lou Reed is not. Why, then, do some of us have this latter experience? In chapter eleven I will answer these related questions.

The second problem is a more general one: if there are many different phenomenal experiences, then it might be necessary to investigate each and every one of these kinds separately, both with respect to the why-aspect of the problem of phenomenal consciousness, as well as with respect to the how-aspect. The consequence of this is that we have to be patient: a solution to these aspects of the problem of phenomenal consciousness cannot be given in a single article or book, it will take many years of investigation before all the answers will be given. In this dissertation I hope to contribute to the answer to the general question *why is there phenomenal consciousness at all?*

Now I have made clear my assumptions, I will first present the *evolutionary shorthand theory* of the phenomenal mind in *chapter two*. This theory is the result of applying the natural method to different phenomenal experiences. The method is one where data from different scientific disciplines and from phenomenology are combined. The successful application of this method, then, can be interpreted as an argument against the pessimistic view that a science of the phenomenal mind is impossible.
CHAPTER TWO
THE EVOLUTIONARY SHORTHAND THEORY OF THE PHENOMENAL MIND

The upshot of this dissertation is to show that phenomenal experiences form a heterogeneous set of causally efficacious internal and direct presentations. In most normal conditions they are used to represent real properties of external objects or bodily states in a nonconceptual, indirect though reliable way, approximately projecting them back on these states with which they causally covary. In other words: they represent or track those properties. The normal conditions are those conditions in which the experiencing organisms evolved. Phenomenal representations are a species-specific evolutionary shorthand for complex information about the environment or the body that − in principle − could also be represented in another less economic way. It is precisely this economic manner of representing, which makes phenomenal consciousness a feature that is selected for in evolution by natural selection. In all other conditions phenomenal experiences are either just presentations not representing anything (e.g. afterimages), or they are used to represent features that are not found in the environment in which the experiencing organisms evolved (e.g. the beauty of art). Since phenomenal experiences were selected for their representational abilities, experiencing organisms have the natural inclination to interpret any phenomenal experience as representational, even if it is merely a presentation.

In the present chapter I will explain what I mean by the terms I use in this characterization. I will also discuss differences of opinion regarding the different aspects of this theory with others who defend a representational theory of the phenomenal mind. As I said before, my intention is not to present a new theory, but to show that the different aspects of this theory are all supported by empirical evidence, which will be presented in the chapters seven through eleven. One aspects of this characterization are already argued for: the claim that phenomenal experiences form a heterogeneous set is demonstrated in chapter one.

This chapter is meant both as a presentation of the theory and as a guide for reading the chapters to come. The different aspects of phenomenal consciousness are discussed, but the main body of evidence will be presented in the chapters seven through eleven. Very little empirical data on which this theory is based will be presented in this chapter. This means that most arguments will be brief. It is not, however, my intention to convince the reader in this chapter of any of the aspects I discuss, only to clarify these different aspects and show how I will argue for them and where I differ from those who say similar things about the phenomenal mind. Evidently, this way of presenting the theory brings with it some repetition of arguments, but the advantage will be that the reader has a good idea what the point of the different case studies will be.
1. phenomenal experiences are presentations

I am not the only one defending a representational account of the phenomenal mind. However, there is a big difference between my theory and those of for instance Michael Tye (1995) and Fred Dretske (1995). They both maintain that all phenomenal experiences are representational states.

In my view, all experiences and feelings have representational content, not just perceptual experiences. (Tye 1995: xv)

All mental facts are representational facts. (Dretske 1995: xiii)

It is here that the evolutionary shorthand theory of the phenomenal mind differs the most from other representational accounts of phenomenal consciousness. I do not agree that every experience or feeling is representational. As I will show in chapter seven, phenomenal experiences are first and for all presentations we have experiences whether they represent something or not. The main idea here is that presentations do not have any intrinsic meaning or purpose. Phenomenal experiences are first and for all presentations, nothing more. The idea is that we have experiences that can be used as representations. Consider the following analogy. A stone can be used as a hammer, but not all stones are hammers. By analogy the same goes for phenomenal experiences: phenomenal experiences can be used as representations, but not all phenomenal experiences are representations. Some phenomenal experiences are just presentations.

This presentational character of phenomenal experiences will be spelled out in more detail in the rest of this dissertation. For the moment it suffices to say that we can and do have nonrepresentational experiences. A good example is an afterimage: even though one experiences a certain color and a certain shape one knows that there is no real object that has this color or shape. I will also return to this example.

The discussion of the differences of opinion with other representationalist theories of the phenomenal mind are discussions within a large framework of agreement. I believe that most of the point of disagreement stem from my acceptance of the presentational character of phenomenal experiences.

2. phenomenal experiences are direct presentations
Phenomenal experiences are *direct* presentations. If one has an experience of red, this is not mediated by anything else: the experiences are immediate presentations. That phenomenal experiences are direct experiences is probably not a controversial claim, though I mention it here to make sure that the direct character of the presentation is distinguished from the indirect character of what is represented by a phenomenal representation, for which I will argue later.

### 3. Phenomenal experiences are internal states

One of the main points of the following chapters is to show that phenomenal experiences are states internal to biological creatures with a certain kind of brain. Phenomenal experiences are located in the head, even though they are experienced *as if* they are out there (with the exception of headaches and the like). I take it that most people would claim that it is the grass that is green, that it is the tomato that is red, that the pain due to hitting your thumb with a hammer is in your thumb, and that it is the face of your loved one which is beautiful. I will argue that every experience is in the head, supervenient on a spatiotemporal brain state. There are no experiences or parts of experiences out there. So, I am explicitly arguing against any *externalist* theory of phenomenal consciousness that places (a part of) phenomenal experiences outside the brain.

If externalism is the thesis that the external world is part of the *cause* of the experiences, then of course I am an externalist: the physical world exists and it causally interacts with us.

Externalism, though, is mostly taken as a thesis about *content*. In this view externalism requires realism about the content of the mental states that is out there.

> [E]xternalism is incompatible with the likes of idealism and phenomenalism and projectivism. In other words, externalism presupposes (a form of) realism about the external world: it takes worldly facts to be fixed mind-independently, so that these can be exploited in the fixing of mental facts. (McGinn 1989: 10)

This means that a secondary quality view of color (and other phenomenal properties) is incompatible with externalism.

> [I]f colours or moral values, say, are taken to be projected onto the world by the mind, then there can be no explaining the relevant distinctions of mental content by reference to prior worldly distinctions. If an object’s being red consists in its looking red (crudely), then it cannot also be true that its looking red consists in being brought about by red things. […] Hence if we have reason to believe externalism (about some class of mental states), then we thereby have reason to favour realism about the subject matter of those states. The truth of externalism about a class of mental contents would entail that the relevant sector of reality is not mentally constituted. (McGinn 1989: 11)
To turn McGinn’s reasoning the other way round: if we have reason not to believe realists about some subject matter, then we thereby have reason to favor internalism about the class of mental states that have this subject matter as content. Hence, the primary quality view of color and externalism about color stand or fall together. In the chapters on synesthetic and color experiences I will argue for the secondary quality view of colors and hence for internalism about the content of phenomenal states.

Of course I do not claim that there are no tomatoes out there of which we say that they are red. A thought about a tomato is about a real tomato out there. What I am an internalist about is a particular kind of content, viz. those that are about the secondary properties of an object. This is compatible with an externalism about the primary properties of objects. If I perceive a red and round tomato, the experience of roundness is brought about by the actual round shape of the tomato. Here we can say that the content of the experiences is the actual roundness of the tomato, which is an externalist view. In the case of the redness, the content is in the head. This comes down to the claim that the phenomenal experience is its own content: we cannot be internalist about the location of the phenomenal experience and externalist about its content, for they are both one and the same. Leaving aside the implications of Tye’s claim that all phenomenal states are representational, he says something very similar: “According to the theory developed in Tye 1995, phenomenal character is one and the same as representational content that meets certain conditions.” (2000: 45) If the phenomenal character itself is content, then one should be either an internalist about character and content or an externalist. Dretske and Tye opt for the latter:

The representations are [in the brain], but their content is not. In this sense, the mind isn’t in the head any more than stories (i.e., story contents) are in books. (Dretske 1995: 38)

On the PANIC theory, as I have elucidated it, phenomenal character ain’t in the head. So the thesis that the phenomenal supervenes on the neural is false. | (Tye 1995: 194)

4. **phenomenal experiences are projected**

The claim that phenomenal experiences or phenomenal content is located in the head usually is met with great resistance and is often regarded as naïve. Is it not the grass that is experienced as being green? I agree that this indeed is how it seems, but that the actual experience of green is in the head and is *projected* onto the grass. This seems to be somewhat counterintuitive, for it seems that if the grass is green and the grass is out there, then the green is out there. However, the fact that the green seems out there does not imply that is actually is out there. This is the mistake both Max Velmans and Michael Tye seem to make, though in different ways.
Velmans (e.g. 1991a; 1991b; 2000) also claims that the right theory of consciousness is one that claims that the phenomenal world is out there, and that it is projected there. One of the problems with his theory is that he is not clear about what is projected and from where. He explicitly denies that phenomenal experiences are located in the head, for in his opinion that would amount to two phenomenal worlds for every person: one in here and one out there. Velmans argues that there is just one phenomenal world and that it is out there. He remains silent about what is projected or what the ontological status of the projection – the phenomenal world – is.

This shows a problem with the notion of projection. Normally, if we use this word, we do so with respect to an apparatus with a light bulb in it emitting photons that travel through a slide, eventually projecting a picture on a screen. In this case something actually changes at the screen, and other people (or machines) can detect this change. However, the projection of experiences onto the cause of the experience does not add anything at all. It is a projection as it is the case in afterimages. If I have been looking at a colored spot on a piece of paper right in front of me, and then look at the white wall ten meters away, I will project an afterimage on this wall. But the wall is not changed, there are no photons coming from my eyes projecting the picture on the wall for others to see. For me however, there seems to be a large colored spot on the wall. This is the way in which I will use the notion of projection. If this is what Velmans means by the claim that the phenomenal world is out there, then I agree. However, I do not think that this is what he means, for he denies that there are phenomenal states inside the head (with the exception of headaches and the like). In chapter eight I will discuss his theory in more detail.

Tye’s claim that the green of the grass is out there is a totally different claim. He bases his view on the transparency of experiences. Suppose you have an experience of a blue square. Now, examine the experience. Tye argues that there does not seem to be more than that what it is an experience of a blue square: “one experiences blue and square as features of an external surface, but introspection does not seem to reveal any further distinctive features of the experience over and above what one experiences[,]” (1995: 30) The same goes for other experiences, like pain. I agree with his following conclusion: “The lesson of the problem of transparency is that phenomenology ain’t in the head. […] To discover what it’s like, you need to look outside the head to what the brain states represent. Phenomenology is, in this way, externally based.” (1995: 151) This claim however can be taken in two ways. My interpretation of the claim that phenomenology is externally based would be that the causes of our experiences are external: we
represent those causes in a certain way, projecting them back, as if the presentation was out there. Tye means something different. He claims that from the fact that experiences are transparent it follows that there really are blue squares out there, independent of us. “[T]he best available hypothesis is that color is an objective, physical property of external things.” (2000: 167) This means that Tye has to reject Locke’s secondary qualities. Of course, he is right that a Lockean account of color experiences includes the thesis that experiences of colors involve an aspect of illusion; that the secondary quality view of colors implies that: “there is a basic illusion involved in normal experiences of color, that colors are really (response-dependent) relational properties even though we experience them as nonrelational. That, it seems to me, is just not credible.” (1995: 145) Believing that something is not credible is not a good argument for rejecting the view that colors are secondary properties, and to claim that: “Colors are objective, physical features of objects and surfaces.” (1995: 150) In chapter ten I will present data that show that this view has to be wrong.

5. **Phenomenal presentations usually are used as representations**

Phenomenal presentations usually are used as representations: the sugar tastes sweet, the grass is green and the music is loud. In chapter eleven I will argue that it is in virtue of this ability of the presentations that they were selected for in evolution by natural selection. From an evolutionary point of view, experiences that are not used as representations are useless to the organism, hence were not selected for. This does not mean that all presentations are used as representations, only that those selected for were. Experiences that were not selected for are – by definition – evolutionary epiphenomena. As I will argue in chapter nine, dreams may well be such phenomenal experiences: they exist as spandrels of sleep and the evolutionary selected capacity of our mind-brain to make sense of our experiences by immediate interpretation of the presentations as representations. In the same chapter I will also argue that the fact that experiences are selected for their usefulness as representations has the consequence that we immediately project any experience onto the objects we are perceiving; we have evolved in such a way that most of our presentations are used as representations, resulting in a natural inclination to interpret almost all experiences as properties of the objects around us. This natural inclination usually goes quite far: we have the natural inclination to interpret experiences as properties of objects even if there are only experiences and no objects, as is the case in dreams, hallucinations and afterimages.
This natural inclination might be the reason why other representationalists claim that all phenomenal experiences are representational, and that "[i]n the case of an afterimage, what is represented is that there is something perpendicular to the line of sight that is \( F \) and \( G \), typically, something that has a dim and hazy character." (Tye 1995: 107) Since there really is no object that corresponds to the experience and Tye claims that every phenomenal experience is a representation, he claims that "an afterimage is a misrepresentation." (1995: 109) In my view this is incorrect. Since we all know that there is no corresponding object to an afterimage, we should say that afterimages are merely presentations. Furthermore, because we evolved in such a way as to immediately interpret any presentation as a representation we are inclined to do this too in the case of an afterimage. But even thought we have this inclination, in this specific case we do not interpret the presentation as a representation, for we know that there is nothing out there that is represented.

Another case where we know that there is nothing corresponding to our experiences is when one has the temporal lobes stimulated by transcranial magnetic stimulation (TMS). Subjects are well aware that their experiences are artificially induced via TMS. Still they often try to make sense of their experiences by interpreting them as religious experiences (Persinger 1987; Hitt 1999). Here we also see that the brain tries to interpret presentations as representations, even if it has to come up with a pretty weird scenario to interpret the experiences in a coherent way.

Using TMS to investigate consciousness is still a new way of gathering data about the phenomenal mind, so there is not much to go on right now. The data provided by Persinger are coherent with the evolutionary shorthand theory of the phenomenal mind, for they show that experiences can occur without the usual stimulus, hence that the experience is a mere presentation. This presentation can be interpreted as a representation, even though one knows that it is not a representation at all.

6. **Phenomenal representations are nonconceptual representations**

For those experiences that are used as representations, I agree with both Tye (1995: 137) and Dretske (1995: 1) that phenomenal experiences have nonconceptual content. According to Tye, the phenomenal content "is content that is appropriately poised for use by the cognitive systems [...]. I call this the PANIC theory of phenomenal character: phenomenal character is one and the same as Poised Abstract Nonconceptual Intentional Content." (1995: 137)
Tye says that the content is abstract, for no particular objects enter into the content. A fire truck is red, just as a tomato is, hence the phenomenal character of red does not involve trucks or tomatoes, but is abstracted away from these objects.

The claim made by Tye and Dretske that phenomenal content is nonconceptual has to be understood as saying that one can have pain without having a higher-order concept – or a metarepresentation as Dretske calls it – of pain.

With this I wholeheartedly agree: content can be nonconceptual. Peter Carruthers (1998) interprets the view that phenomenal content is nonconceptual as a defense of a first-order representational (FOR) theory. However, this is incorrect, for it is possible to have higher-order nonconceptual content.

First of all a FOR theory in my opinion is incorrect, for not all phenomenal experiences are representational. Secondly, the usual interpretation of higher-order mental states as conceptual states is incorrect, though I agree that in most cases phenomenal states are indeed used as first-order representations. But this is not always the case. Intuitively this makes sense by considering the following. A photo of self-portrait of Rembrandt is a representation of a representation of Rembrandt, hence a higher-order representation. But just as the painting is a nonconceptual representation, so is the photo. Hence, higher-order representations can be nonconceptual. As will become clear in the discussion of synesthesia some phenomenal states are naturally used to transparently represent other representational mental states, which in effect makes them higher-order phenomenal states. If a phenomenal presentation is used in this way, it still is a nonconceptual representation.

The first reason that higher-order representationalist (HOR) theories of phenomenal consciousness are also incorrect is the same as why FOR theories are incorrect: not every phenomenal experience is a representational state. Furthermore, it denies any creature lacking higher-order representations (interpreted here as concepts of conceptual or nonconceptual representations) phenomenal experiences, which is absurd. One might bit the bullet on this, as Peter Carruthers does, agreeing that: “HOR-theories must deny phenomenal-consciousness to
the mental states of animals (and very young children)]” (1998: 214) I think this is obviously wrong: a baby cries when it is hungry and smiles when it sees a familiar face, which is typical behavior expressing the corresponding phenomenal experiences. Also, in the case of animals, a dog seeing another dog, his boss, or about to get his food seems very happy. This, in my opinion, refutes Carruthers’ HOR theory beyond any doubt, even though he does not accept this as an argument (1998: 214).

So, in essence I disagree with both FOR and HOR theories, for both theories take all higher-order representations as conceptual representations. Whereas HOR theories are wrong, for an organism might have phenomenal experiences without higher-order representations, FOR theories seem to be the victim of a conceptual interpretation of the notion of a higher-order representation, thereby leaving out the higher-order nonconceptual phenomenal representations. Phenomenal presentations, if used as representations, are nonconceptual representations that are either first-order or second-order representations.

A more serious problem with Tye’s theory is that it does seem to lose the phenomenal character of consciousness along the way. It seems that the only difference between propositional attitudes and phenomenal experiences is that the first are conceptual and the second are nonconceptual. But being nonconceptual does not account for being phenomenal. A photo of a tree or the rings of a tree are also nonconceptual representations, but without phenomenology. So the PANIC theory seems to leave out an essential feature of conscious experiences.

7. **phenomenal representations are indirect representations**

In a theory of representations we have to make various distinctions. First, we need to distinguish between natural and nonnatural representations. The nonnatural representations include pictures, movies, words, computerbits, etcetera.

Dretske distinguishes natural from conventional representations.

When a thing’s informational functions are derived from the intentions and purposes of its designers, builders and users [..], I call the resulting representations *conventional*. Representations that are not conventional are *natural*. (Dretske 1995: 7)

I think we can equate the nonnatural with the conventional representations, though there may seem to be a problem with the language that is mostly used in the discussion of evolution. We often speak of something that is designed by evolution or the functions of certain biological
features. Of course, it is always added that this language is metaphorical: evolution has no intentions. So, if we drop the literal interpretation of such language, there is no danger of interpreting the phenomenal presentations that were selected for in evolution as conventional representations.

The nonnatural representations are not the ones we are interested in, since phenomenal representations are natural representations: representations of a biological system. These biological systems represent features of their system or their environment in a manner relative to the creature and its interests. We represent a mountain as big, because it is bigger than we are; we represent something as edible, because it is edible for us; we represent some people as beautiful, because they are beautiful to us.

A further distinction that should be made is that between direct and indirect representations. Some property $A$ is represented directly if it is represented as $A$, and indirectly when it is represented – but not misrepresented – as $B$. These two ways of representing are the representational counterparts of primary and secondary properties as distinguished by John Locke (1690).

At first sight there is a certain risk involved in presenting a theory in which the representations of a biological system are all relative to the interest of the creature: all represented properties seem to become secondary qualities. A wall has a certain shape and color for us: if we were as small as an electron the wall would be a vast colorless space, with few other particles in it. Not only for colors but also for shapes to exist an observer with a particular configuration is needed. However, this is not a serious problem. Even if it is true that a red wall is hard for us, because it resists our pushing and we cannot pass through it, representing it as hard is a direct representation of a property the wall has independently of us, hence it is a representation of a primary property. It is an objective feature of walls, that objects that belong to the same macrophysical domain cannot be at the very same place as the wall or pass through it. Humans are one kind of macrophysical entities that fit this category. Entities of other domains might be either so small that a wall is just a vast space with occasionally another microphysical entity, or so big that walls are no obstacles at all. But this does not mean that the solidity, hardness and shape of the wall depend on us for their existence: they exist independently of us, but these properties cannot be encountered at higher or lower levels in the mereological model of the world. So, shape, solidity, liquidity, hardness and the like are all primary qualities of objects. If humans represent walls as being solid, they represent this property directly: property $A$ is represented as $A$. In such a representation the level
in the mereological model is implicit. If I represent a wall as being hard, I represent a property of a macrophysical object. The fact that on a different level the object is a vast space through which microphysical particles can pass is irrelevant for the representation at the macrophysical level.

Before I turn to the representation of secondary properties, I have to say something about the form of the representation. To represent a wall as square or hard does not imply that the representation itself is square and hard. Just stating in words that “the wall is square and hard” is enough to represent these properties of the wall directly.

The color of the wall on the other hand is a secondary property: in order for the color red to exist there have to be creatures that respond to certain primary properties in the way that they experience red.

If we take a look at the parts of our universe where there are no observing creatures, then we can safely state that phenomenal experiences of colors are absent there. Logically, this does not imply that colors themselves are absent. Just as hardness still exists even when it is not observed, colors might exist, even though they are not experienced. In chapter ten I will demonstrate that colors do not exist: there are only experiences of colors. This seems strange at first sight, for how can we have experiences of colors, when there are no colors? I believe the answer to be relatively simple: we represent some real properties in an indirect way, viz. something is A independent from us, but we represent it as B (as having a certain color for instance). It is precisely this strange way of representing that makes phenomenal representations so special. This also has the consequence that talk about colors is as-if talk, even though most of us do not realize this.

Another useful way to distinguish between direct and indirect representations is with respect to the physical laws that hold on a certain level – or between levels – of the mereological model. Physical laws describe the behavior of physical entities at the appropriate level of the mereological model. These laws are independent of human beings or other observers. Hence, everything that is referred to by the terms in a physical law, such as mass and motion, can exist without observers. So, mass, motion, shape, hardness, etcetera all are primary properties of objects in our universe, since they can function in physical laws that exist independently of us. Colors, smells, sounds, tastes do not feature in any physical law, only their various objective physical counterparts – like wavelengths and chemical composition – do. This means that direct representations can feature in physical laws while indirect representations cannot. I should note
that this cannot be take as an argument against physicalism, for it will become clear that what is indirectly represented can figure in physical laws too, so nothing is left out by broad physicalism by not using certain representations.

8. phenomenal representations represent environmental and bodily properties
Phenomenal representations are representations of a certain kind, they also have species-specific contents: they can either be used to represent properties of the environment of the representing system or of the bodily states of this system. Here I agree with Tye, who claims that sensory representations are representations of “either the outside world or of certain sorts of internal, bodily changes.” (1995: 94)

As I already said, phenomenal experiences are inside the head, but their content seems to be out there: it is the grass that seems to be green. Whether the grass actually is green, does not matter for the claim that the grass is represented as such. The same goes for the pain in the thumb: whether or not the thumb actually is the location of the pain is irrelevant for the truth of the statement that the thumb is represented as painful. What we can say at least is that there is some property of grass that is represented as green, and some property of my thumb that is represented as painful, hence that phenomenal experiences represent environmental or bodily properties. They do this in an indirect manner. In chapter eleven I will discuss the experiences of beauty evoked by faces to show that the feeling of beauty is a presentation that is projected back onto its causal origin (a face) and that indirectly represents primary properties of the person who’s face it is the observer looks at. These primary properties are fertility or health.

9. phenomenal representations are approximately projected to the cause of the experience
I already said a few times that the experiences we have are projected onto the cause of the experiences. The tree is the source of our green-experience, and we perceive the tree to be green. So the green-experience is projected onto the tree. However, it should be noted that the projected location of the phenomenal world and the actual physical world do not map neatly. Variances in depth estimations between individuals shows that the projected locations phenomenal worlds of each of us differ, hence cannot all be precisely mapped onto the physical world (Koenderink et al. 1995; Koenderink et al. 1997). However, even though the depth perception is not veridical and differs from individual to individual, there are no problems in
dealing with the world, for the mismatch between true depth and experienced depth applies throughout the visual field (1997: 835).

10. phenomenal representations represent in a reliability way: causal covariance

The distinction between direct and indirect representations brings us to the question of reliability. If a feature $A$ is represented as $A$, this evidently is a reliable representation. But what about the representation of $B$ as $C$? Here, the usual indirect representations are reliable, in the sense that the same features of the environment usually are represented in the same indirect way: the grass is normally represented as green, ripe tomatoes as red, and sugar as sweet. The primary properties of sugar have causal impact on our taste-buds, that send signals to the CNS, leading us to experiences of sweetness. Most of the times we eat sugar this occurs; there is a causal covariance between the eating of sugar and the experience of sweetness. The same goes for all other phenomenal representations: they are caused by stimuli, and the same stimulus normally leads to the same experience. This is what Michael Tye calls “tracking” (1995: 101): the experience of sweetness tracks real features of the sugar cube. The conditions for successful representation have to be optimal (1995: 101). I think the term “optimal” is too strong here: in optimal viewing conditions a banana seems to be yellow, but this is still the case at dusk, when the viewing conditions are no longer optimal. This phenomenon is called color constancy. I think it is better to use the notion “normal” for it is less restrictive then optimal, allowing for correct tracking of properties without the optimal conditions being obtained. The conditions of the environment are normal when they are the same as the conditions in which the representing system evolved. So the viewing conditions of our cities at night are far from normal. Also, the representing system should be in normal condition, which means that it should not be damaged. Here we probably can just look at what the most common features are, which for human beings include having two eyes with three kinds of cones. If the conditions are not normal, then misrepresentation might occur.

If the light is dim, or if the context of a normal external stimulus is changed (as I will show in chapter ten using the Land color Mondrian experiment) the phenomenal representation does not successfully represent the properties of an object any more; e.g. a banana might no longer seem yellow.

As I said before, the main difference between the evolutionary shorthand theory and other representational theories of phenomenal consciousness is that I believe that phenomenal
experiences can be used as representations, but are not always representational as Tye and Dretske argue. This has its consequences for the view on afterimages. Tye and Dretske have to argue that afterimages are representational, and that since there is no object corresponding to the experiences, they are misrepresentations. Dretske makes the claim that there are two kinds of misrepresentations. First there is something that is incorrectly represented (e.g. a cat is represented as a rabbit). Secondly, there is not something that is represented as an object. “If there is no object of representation, then S represents (i.e., misrepresents) this as being blue when there is no this.” (1995: 27) Clearly, an afterimage would be a good example of the second kind of misrepresentation.

As we saw in section five Tye also regards afterimages as misrepresentations (1995: 109). However, I do not see how the conditions are not normal (or even optimal) here. If one of our perfectly healthy ancestors looked briefly into the sun he would have experienced afterimages afterwards. Both the environment and the organism were in normal condition, still there was an afterimage. If misrepresentations only occur if normal conditions do not obtain, afterimages cannot be misrepresentations. Afterimages are merely presentations. That is what I believe is the correct view. (I will discuss misrepresentation in more detail in chapter eight.)

In the case of afterimages the conditions are normal, but these phenomenal experiences are not representational. If the conditions are not normal, misrepresentations might occur. In other words: if the conditions are not normal, then the representation does not track what one believes it does: If the light is dark, a rat might be represented as a cat. The representation in this case would not be tracking rats, for the object of the representation would behave differently from the instances of the representation of which it would be a representation of under normal conditions. (Note that interpreting a presentation as a representation of something in the absence of anything, is different from a representation of something that no longer exists. A picture of my grandmother is still a representation of my grandmother, even though she is no longer alive: a representation obviously can survive the represented.)

For a phenomenal presentation to be a reliable representation, then, normal conditions have to obtain, and the presentation has to be used as a representation. For indirect representation a problem similar to that of misrepresentation seems to rise, for in both cases A is represented as B. In the case of misrepresentation a cat might be represented as a rat, and in the case of phenomenal representation – according to the evolutionary shorthand theory – no thing that is
indirectly represented as being B ever is B (e.g. no things that are represented as green are green). I will turn to this problem now.

II. evolution’s shorthand

Does it make evolutionary sense that we represent our environment in an indirect way? Why represent something as something different from what it is – even if it is a reliable representation, that is, even if it occurs time and again with the same stimulus?

Here the answer is that evolution does not care about whether a representation is direct or indirect, it only cares about what is fitness enhancing. In the case of phenomenal representations, how things are represented could also have been represented in another – viz. direct – way (of course the phenomenal feel would be lacking). Why, then, did we evolve having those puzzling representations? The answer is simple: phenomenal representations are evolution’s shorthand: the indirect way of representing is much shorter and faster than the direct way would be. This is especially useful in life-threatening situations, but also in more mundane situations as looking for food. Note that presentations are always direct, and that these direct presentations are used as indirect representations. It is precisely the immediate character of the presentation that gives the indirect way of the phenomenal representations the advantage over the direct way of nonphenomenal representations.

Suppose that the natural habitat of Homo is the Savannah of eastern Africa. Even though they had the ability to make weapons, the first weapons were very primitive and predators were a threat to them. It makes good evolutionary sense to be able to pick up information about the environment very quickly. Picking up affordances has to be done in an economic manner. Being afraid if one sees a predator just is one of the phenomenal experiences that precisely does this job. When we see a predator – say a lion – we immediately are scared and will stand still, try to hide, flee or attack. The information could also be specified along the following lines: “There is a big feline animal, which I know to be carnivorous and seems to be hungry, probably intending to have me as his dinner. This is something which is very bad for my gene-survival. Hence I should either hide, flee, or try to render the attacker powerless.” By the time the prey is halfway through this reasoning, he will be the lion’s dinner. Hence, nonphenomenal representations can indeed express the same content – from an external observer’s point of view – but creatures that would do this could not have been evolved, for predators would get to them.
An additional problem would be that the creatures have to have some knowledge of neodarwinistic evolution theory, which evidently most creatures lack. It is physically impossible that a creature could have evolved based on knowing how evolution via natural selection works, for the very first creature that should do this would already have to have this knowledge. So, if a certain kind of knowledge was fitness enhancing, it had to be another kind, a kind in which explicit knowledge about the mechanisms of evolution were absent, but which works just as well – or even better: phenomenal experiences.

This means that phenomenal experiences – just like linguistic thoughts – create intensional contexts. If Ralph believes that the man in the brown coat is a spy, and that man happens to be Ortcutt, then it is not necessarily true that Ralph believes that Ortcutt is a spy, for he might not know that the man in the brown coat is Ortcutt. The same goes for phenomenal representations. Even though being scared of a predator is a representation that is – from an objective point of view – equivalent to a very long nonphenomenal representation of which knowledge of the neodarwinistic theory of evolution is expressed in combination with an evaluation of the situation, we cannot substitute the latter representation for the first. This also shows why phenomenal representations cannot be reduced to other representations: they create intensional contexts.

I call the above theory the *evolutionary shorthand theory of the phenomenal mind*. Its main claim is that phenomenal experiences are presentations that can be used as representations, thereby becoming candidates for evolutionary selection. This main claim and the other claims of this theory are all the result of applying the natural method to a variety of phenomenal experiences. These phenomenal experiences are discussed in the chapters seven through eleven. Before I turn to these case studies, I will first address several views that would imply that a scientific study of phenomenal consciousness within a broadly physical framework is not possible, either in principle, or at this moment. I will start by discussing both the usual argument for dualism (chapter three) and this position itself (chapter four).
If it is accepted that science assumes monism, then the biggest threat to a scientific solution of the problem of phenomenal consciousness evidently comes from dualism. In this and the next chapter I will argue against this position using a two-fold strategy. In the present chapter I will show that the most frequently used argument for dualism is invalid, and in chapter four I will – for the sake of argument – accept the most recent version of property dualism showing that this leads to a *reductio ad absurdum*. In both chapters zombies play a central role. In this chapter I will show that the argument in favor of zombies is invalid, and in the next that the notion of a zombie itself is incoherent.

1. **how dualism threatens a scientific solution to the problem of phenomenal consciousness**

In the contemporary consciousness debate almost no one is a substance dualist. Most dualists are property dualists, where the physical world has primacy over the phenomenal world. As I already said in chapter one, we should distinguish between two kinds of dualism. First, there is the kind that is compatible with neutral monism: broad physicalism. The world is made out of one kind of stuff, but there are different ontological levels. We can be dualists about liquidity in this manner: at the level of individual molecules there is no such thing as liquidity, this is a real property of a higher level: groups of molecules. This kind of dualism might better be called pluralism, for there are more than two levels in the mereological model.

Secondly, there is the tough kind of dualism that threatens a scientific solution to the problem of phenomenal consciousness that makes a much stronger claim. According to this much stronger kind of property dualism, the phenomenal is totally different in kind from anything that might be called broadly physical, even though the physical is necessary for the phenomenal. The most recent version of this view is the position of David Chalmers (1996b) who argues that a physical duplicate of our world is logically possible (hence sharing all the broadly physical properties of our world) without there being any phenomenal property in that world. In this so-called zombie world certain natural (nonphysical) laws are absent, and it is because this world lacks these laws that phenomenal consciousness does not arise. In this zombie world there are biological creatures that are our physical counterparts – our zombie twins so to speak – that are physically identical and functionally equivalent to us, but that do not have any phenomenal experience whatsoever.
There is nothing it is like to be a zombie. In this and the following chapter I will discuss this tough kind of dualism.

Usually the kind of argument given for dualism is a conceivability argument. This starts with the claim that zombies (or otherwise phenomenally different physical duplicates) are conceivable. In addition, there is supposed to be some route from the conceivability to the possibility of the zombies, thereby proving dualism. The further implication is that – since science assumes monism – there can never be a scientific solution to the problem of phenomenal consciousness.

In this chapter I will argue that conceivability arguments are not valid. This severely weakens the case for tough dualism, for this leaves its proponents empty-handed unless they come up with another argument. I will first present the traditional framework in which different kinds of possibility are distinguished. I then will point out the kind of possibility that is needed for conceivability arguments to show that dualism is false, and see whether the other kinds of possibility provide a route to it. After this, I will present Chalmers’ main conceivability argument. Lastly I will evaluate whether this argument indeed provides a way to defend dualism. I will argue that it does not.

2. possibility

Traditionally a variety of possibilities has been distinguished. I will describe the most important varieties, namely logical, epistemic, ontological, and metaphysical possibility.

2.1 logical possibility

In general, we can say that a statement \( s \) is logically possible, when it does not contain a (conceptual) contradiction; when it is not self-contradictory.

If, for example, I were to say that it is possible that beginning tomorrow the temperature of physical objects will vary with their colors, or that it is possible that in one minute the desk on which I am writing will vanish from sight, one thing which I should mean is that the statements “Beginning tomorrow the temperature of physical objects will vary with their colors” and “In one minute this desk will vanish from sight” are not self-contradictory. (Malcolm 1950: 252-253)

Another characterization comes from Jon Barwise and John Etchemendy:

Intuitively, a sentence is logically possible if it could be (or could have been) true, at least on logical grounds. There might be some other reason, say physical, why the statement could not be true, but there are no logical reasons preventing it. For example, it is not physically possible to go faster than the speed of light, though it is logically possible: they do it on Star Trek all the time. On the other
hand, it is not even logically possible for an object not to be identical to itself. That would simply violate the meaning of identity. (1999: 94)

There are at least two accounts of logical possibility: what is logically possible on one account might not be logically possible on another. I think it is best to make a subdivision within logical possibility. Statements can contain a contradiction on formal (syntactical) or on conceptual (semantical) grounds. Something is formally or syntactically possible when it is not of the form \( (A \land \neg A) \). This means that only explicit contradictions fail the test. So it is formally impossible that the morning star is not the morning star, while it is formally possible that Kennedy survived his assassination since, even though the statement is false because of the concepts used, it is not of the form \( (A \land \neg A) \). The evaluation of this last example changes if one uses the conceptual variety of logical possibility.

"The conceptual criterion marks possibility by conceptual content. It says a claim can be true when its falsity is not ensured by the concepts from which it's built." (Sturgeon 2000: 103) Chalmers also uses logical possibility in this conceptual variety.

In determining whether it is logically possible that some statement is true, the constraints are largely conceptual. The notion of a male vixen is contradictory, so a male vixen is logically impossible; the notion of a flying telephone is conceptually coherent, if a little out of the ordinary, so a flying telephone is logically possible. (1996b: 35)

What we see in this subdivision is that there is nesting of possibilities: that which is conceptually possible is also formally possible, but not vice versa: conceptual possibility is nested in formal possibility.

2.2 epistemic possibility

The second kind of possibility is epistemic possibility.

A state of affairs is deemed epistemically possible if it is not ruled out by the knowledge, both a priori and scientific, that we happen to have at a certain time. In other words, if, relative to the knowledge we happen to have at a certain time, it would be inconsistent to admit a certain thing, event, or state of affairs as actual, then the existence of that thing, event, or state of affairs is epistemically impossible for us at that time. It is clear that because the state of our knowledge is constantly changing, what is epistemically possible at one time may be declared epistemically impossible at another time. (Aune 1967: 423)

Here we should distinguish between what we might call mere and positive epistemic possibility. In the first case there is no epistemic reason to believe something is not possible. For instance, we have no reason not to believe that black holes are passways to other dimensions, so we might safely conclude that it is merely epistemically possible that black holes are passways to other
dimensions. In other words: “When an agent claims that something is possible, he or she often means that it is “possible for all I know.” (Barwise 1997: 495)

In the case of positive epistemic possibility, someone has some epistemic reason to believe that something is possible. Or, as Malcolm puts it: “When it is said in ordinary life that “It is possible that so and so will happen,” what is very frequently meant is that there is some reason to believe that so-and-so will happen.” (1950: 254) We might for example have some evidence that Plato was not the author of *The Republic*.

If a Greek scholar were to remark “It’s possible that Plato was not the author of *The Republic*” we should ask “Why do you say so?” and it would be only a joke if he were to reply “The statement ‘Plato was not the author of *The Republic*’ is not self-contradictory [.]” [...] We should naturally interpret his first remark to mean “There is evidence that Plato was not the author of *The Republic*”; and our question “Why do you say so?” would be naturally understood as a request for him to say what the evidence was. (Malcolm 1950: 255)

In this light we might say that is positively epistemically possible that tomorrow biologists might discover giant living insects in the Amazon rain forest. We have some reasons to believe this, for we know from the fossil record that once there were giant insects on the Earth: we have found their remains. So, once they existed, which is reason enough to believe that they still might be around, just like other animals like crocodiles that have been roaming the Earth for a very long time.

Both mere and positive epistemic possibility differ from logical possibility. Both varieties of logical possibility depend on logic, which is independent from us. We might not see a logical contradiction is a complex statement. If we have no epistemic reason not to believe that the scenario described by that statement is impossible, we might conclude that it is merely epistemically possible.

Positive epistemic possibility also differs from logical possibility, for it allows for degrees: you need reasons to believe that is positively epistemically possible, and these reasons can vary from very weak to very strong. This is not the case with logical possibility: “There can be more or less evidence for a statement, the reasons for believing it can be more or less strong, but a statement cannot be more or less self-contradictory.” (Malcolm 1950: 256)
Since knowledge states can change, something that once was merely epistemically possible can become known to be positively epistemically possible if we find some supporting evidence, or it can become positively epistemically impossible if we find some evidence against it.

### 2.3 ontological possibility

Traditionally logical and epistemic possibility are distinguished from ontological possibility. Ontological possibility, in general, can be characterized as follows. Something is ontologically possible if it does not violate the laws of our world. Here, then we have a kind of possibility of which the actual world in which we live is the judge: whether something is ontologically possible in our world or not, depends on how our world is and on what the laws in our world are. We might distinguish between two kinds of ontological possibility: physical and natural possibility. The first kind of possibility depends on the physical laws, the second on both physical and nonphysical laws.

#### Physical possibility

If there are only physical laws in our world then ontological possibility is physical possibility. “When we ask whether something is physically possible, we are relativizing to the set of physical laws. [...] It does not have to be a known physical law.” (Barwise 1997: 496)

In explaining the notion of positive epistemic possibility I used the example of biologists discovering giant living insects as an example. But this scenario can also be used to clarify physical possibility, for it is not physically possible. Biologists have discovered that nowadays insects cannot become very big in our world, for their respiratory system would not be sufficient to transport oxygen to all their organs. As Barwise notices, we can combine the physical and the epistemic notion of possibility; certain physical laws are known. This makes that now it is both epistemically and physically impossible for insects to be gigantic. In a society where the relevant laws are not known, it still is physically impossible for an insect to be gigantic, but it would still be epistemically possible for all they know. We might put it this way: sometimes we have knowledge that a scenario is ontologically impossible. We then have conclusive reason to claim that

---

1 “The relation between the oxygen consumption of an insect and its tracheal dimensions is given by the formula: $S = k(p - p')A/L$ where $S$ = ml of oxygen used per second; $p$ = partial pressure of oxygen in the atmosphere [...]; $p'$ = partial pressure of oxygen at the ends of the tracheae; $A$ = mean cross-sectional area of the tracheae in sq. cm, $L$ = mean tracheal length in cm and $k$ = the diffusion constant for oxygen[.] [...] Reference to the diffusion formula cited above will show that in non-ventilating insects the maximum possible body size is limited, since oxygen consumption is affected more by a linear increase in size than is the rate of diffusion. The same effects also operates, though to a lesser extend, in ventilating forms since oxygen passes along the finer branches only by diffusion.” (Richards & Davis 1997: 222-223)
the scenario is positively epistemically impossible. In other situations we might have knowledge that indicates that a scenario is ontologically impossible. If that is the case, then we have probable reason to claim that the scenario is positively epistemically possible.

At t = 0 (before the biologists made their discovery) the evidence supporting the positive epistemic possibility was based on the fossil record that contains the remains of giant insects. After the discovery (t = 1) this positive epistemic possibility is nullified, leaving the fossils of giant insects as a problem to be explained. The explanation is simple: when these insects lived, the atmosphere contained more oxygen, resulting in the fact that the air in the tracheae could travel a longer distance before being depleted of oxygen, hence insects could be big.

**Natural possibility**

If there are both physical and nonphysical laws in our world then ontological possibility is *natural possibility*. This view is endorsed by Chalmers. Taken together the physical and nonphysical laws are the natural laws. “A natural possible situation is one that could actually occur in nature, without violating any natural laws.” (Chalmers 1996b: 36) So a natural possibility is a possibility that is allowed by both the set of physical laws and the set of nonphysical laws.

Suppose that in our world nonphysical (hence causally inefficacious) ectoplasmatic entities exist. Say that they do so in virtue of some natural, nonphysical, law that describes their dependence upon some broadly physical feature of our world. How should we evaluate the statement “there are no ectoplasmatic entities in this world”? First, this scenario is a physical possibility, for the absence (or presence) of ectoplasmatic entities does not violate any physical law. This is a somewhat counterintuitive conclusion, for the notion of physical possibility is intuitively taken to cover only the domain of the physical. However, if physical possibility is defined by what is allowed by the physical laws, then anything that falls outside the domain is physically possible, for no nonphysical property can ever violate a physical law (for if it could it apparently was not nonphysical).

Secondly, the hypothesis that there are no ectoplasmatic entities does violate some nonphysical law, viz. the laws that requires that ectoplasmatic entities exist. This makes the scenario naturally impossible, for the nonphysical laws – again *ex hypothesi* – do not allow for their absence given the way the world is now.
It should be noted that epistemic possibility is not nested within ontological possibility, for something might be ontologically possible, while we think we have evidence that it is impossible. Humans are fallible creatures, and our knowledge states may not always reflect reality successfully.

2.4 Metaphysical possibility

The last kind of possibility I will discuss is metaphysical possibility. What is metaphysically possible is not restricted by the laws of our world, but by metaphysical laws (Barwise 1997: 496). “The Metaphysical criterion marks possibility by essence. Its says a claim can be true when its falsity is not ensured by the essence of its truthmakers.” (Sturgeon 2000: 103) Sturgeon argues that the statement ‘David Lewis jumps Mount Everest in a single bound’ is metaphysically possible, while ‘gold is uncomposed’ is not, for gold essentially is composed. The big problem is that it is not always clear what is metaphysically possible and what not. Is the statement that light could have traveled at another speed than it does in our world metaphysically possible or is this speed an essential property of light? I am inclined to think that metaphysical possibility coincides with physical possibility, for I believe that – for instance – the speed light has in our world is an essential property of light. If this is not so one could as well claim that gold is only composed in our universe and not in another metaphysically possible world.

Since there does not seem to be a conceptual contradiction in the notion of a different speed of light, one might be encouraged to equate logical with metaphysical possibility, as Chalmers does (1996b: 42). I think this is not a good idea, for it confuses the criteria by which we evaluate something to be logically or metaphysically possible. Furthermore, the fact that we are ignorant about many essences implies that certain scenarios appear to be possible on conceptual grounds, while in fact they are metaphysically impossible. So, equating metaphysical with conceptual possibility is a bad idea.

3. Dualists should claim that zombies are physically possible

Having distinguished these kinds of possibility, we need to answer the following question. If dualists want to show that physicalism is false by showing that zombies are possible, which of the different kinds of possibility would do just that? Dualists make an ontological claim about our world: there are (broadly) physical properties and there are nonphysical properties, like our conscious experiences. So, dualism is not a claim about logic or epistemology. Hence, dualists have to prove that zombies are ontologically possible. Furthermore, if zombies are used to show
that physicalism is false, then they have to be physically possible, for physicalists endorse the
claim that zombies would violate physical laws, not nonphysical laws. The starting position is one
of ignorance: we do not know whether zombies are physically possible or not, and a conceivability
argument is used to gain the knowledge about the physical possibility.

This also follows from our ordinary talk about possibility. Normally, our claims about what is
possible or not are intended to express what on logical or epistemical grounds we believe to be
ontologically possible. Statements like "the morning star is the evening star and is not the evening
star" and “Kennedy survived his assassination” are useful in clarifying the different kinds of
possibility we can distinguish, but are not normal sentences we use to express our beliefs about
what is possible. In the case of the giant insects we hold at t = 0 that giant insects are positively
epistemically possible, while at t = 1, we nullify this and know that nowadays giant insects are not
ontologically possible in our world. We nullify the epistemic possibility that there can nowadays
be giant insects in our world, because we gained knowledge about the ontology of our world.

In a similar way we can say that at t = 0 (now) we believe that zombies are logically or merely
epistemically possible. We do not see any logical or conceptual contradiction in the claim
“zombies might exist” or have any evidence that shows that they are physically impossible (i.e. we
have no positive epistemic evidence). This belief in the physically possibility of zombies is based
on the conceivability of zombies: conceivability is dependent on the knowledge and insights of the
one who believes the scenario is possible. Now the question arises whether conceivability
provides a route to physical possibility (or physical impossibility). In other words: can we gain
knowledge about physical possibilities using conceivability? I will argue that we cannot.

The main argument against conceivability arguments is that they make the step from either logical
or epistemic possibility to physical possibility, without any argument. Hence an additional
argument is needed, for it is clear that neither of these possibilities provide a route to physical
possibility. It is logically possible (conceivable on logical and conceptual grounds) that black holes
are passways to other dimensions, but we need more than this to be able to confirm the claim
that this is a physical possibility. For the moment we cannot decide on this issue: we are ignorant
and logic or conceptual analysis cannot relief us from this situation. This is true in general: we
cannot gain any new knowledge about the world around us using logic or conceptual analysis. We
need to look at the world itself. Logical possibility, then, is no route to knowledge about physical
possibility. (Logical impossibility of course is a route to physical impossibility.) What about epistemic possibility?

At t = 0 it was positively epistemically possible that giant insects might have been discovered the day before in the Amazon rain forest, while at t = 1 it became clear that this is physically impossible, thereby nullifying the positive epistemic possibility. Positive epistemic possibility is not enough then, for having some reason to believe that A, does not prove that A indeed is the case. Mere epistemic possibility does not lead to new knowledge either: if we have no reason to claim that A is impossible, we can hold A to be merely epistemically possible. But then we are ignorant about the physical possibility of A, and the mere epistemic possibility does not change that.

So, neither logical, nor epistemic possibility provides a route to physical possibility. What about the other two kinds of possibility? The conceivability of zombies then has to be used to establish the natural or metaphysical possibility of zombies, after which an argument is needed to show that from the natural or metaphysical possibility follows that zombies are physically possible.

It is clear that natural possibility is a route to physical possibility, for if no natural laws are violated, no physical laws are violated. Proving that something is naturally possible, involves showing that neither kind of natural law is broken. In the case of the ectoplasmatic entities, we saw that the nonphysical entities evidently cannot violate any physical law for physical laws do not apply to anything nonphysical. Hence anything that is not physical, is physically possible. The case of the zombies is different, though, for dualist cannot start by saying that phenomenal consciousness is nonphysical, hence zombies are physically possible, for that would be begging the question against the physicalist. To provide an argument against materialism, the dualist has to show that zombies are physically possible, i.e. that they do not violate any physical law, for physicalists exactly claim that zombies are physically impossible, because they would violate physical laws.

To be sure, Chalmers claims that it is unlikely that zombies are naturally possible in our world (1996b: 96) for in our world there are nonphysical laws that determine that creatures with our physical make-up have phenomenal experiences.
So, logical, epistemic and natural possibility are nonstarters for an argument against monism. This leaves metaphysical possibility. Could metaphysical possibility be a route to physical possibility? Here too, the answer is no. Granted that it is metaphysically possible for the speed of light to be different than it is in our world (which I deem to be highly unlikely), it still is physically impossible, for that would violate our physical laws. So, metaphysical possibility is no guide to physical possibility either.

In sum, none of the other kinds of possibility provides the dualist with a route to physical possibility. This means that in an argument that starts with conceivability and ends with the rejection of monism, if it makes use of any kind of possibility other than physical possibility, there has to be a step in the argument that shows how this other kind of possibility leads to physical possibility. If this step is made without any additional argument, the argument is flawed.

4. evaluating the conceivability argument

4.1 Chalmers’ main conceivability argument

David Chalmers asks us to consider another universe that is molecule for molecule identical to our world (1996b: 94). In that world no one has phenomenal experiences, everyone there is a phenomenal zombie. This zombie world – according to Chalmers – is logically possible. How does he argue for this claim? He does not. He merely states that it is hard to argue for the logical possibility of anything, and that the only thing in favor of it is the lack of a conceptual contradiction in the description of the scenario under investigation. Since zombies are conceivable – the description of zombies does not contain a conceptual contradiction – zombies are logically possible. “In general, a certain burden of proof lies on those who claim that a given description is logically impossible.” (1996: 96) Assuming that zombies are logically possible, his argument for dualism continues as follows.

4.2 In our world, there are conscious experiences.
4.3 There is a logically possible world physically identical to ours, in which the positive facts about consciousness in our world do not hold.
4.4 Therefore, facts about consciousness are further facts about our world, over and above the physical facts.
4.5 So materialism is false. (1996: 123)

4.2 evaluating the zombie argument

Here we see a clear example of how logical possibility is used to provide a step in the route from the conceivability of zombies to the rejection of materialism. This is not enough, for there has to be an additional argument that the logical possibility implies the physical possibility of zombies. As
we have seen logical possibility alone cannot do the job, not even if it is interpreted as metaphysical possibility. Hence, step three in the above argument is unargued for.

4.3 ignorance

Chalmers’ zombie argument does not do what it is designed to do, viz. prove dualism. The problem lies in our capacity to conceive certain scenarios. If we say that certain scenarios are conceivable, then we believe that they are possible in some sense. Or as Stephen Yablo puts it: “In slogan form: conceiving involves the appearance of possibility.” (1993: 5)

This implies that conceivability is relative to the kind of possibility one has in mind. As a tool to gain knowledge, then, conceivability is of no value, for it merely describes in other terms what one already believes about the possibility of the scenario. If someone claims that zombies are conceivable, he is describing his belief that they are logically or merely epistemically possible, which does not provide us with any new facts. So, an additional argument is needed to show that zombies are physically possible. And this is where we get to the heart of the matter: we are ignorant about whether zombies are physically possible or not, hence there is (at this moment) no additional argument. But nothing interesting follows from ignorance.

In sum, conceivability just redescribes what one already believes about the possibility of a certain scenario. If one is ignorant about the physical possibility of a scenario, conceivability will not help, for that expresses only that one believes that the scenario is possible in another sense than physical possibility. I have shown that such other possibility never is a guide to physical possibility. Hence, conceivability as an argument fails to prove the physical possibility of zombies. This means that dualism has not been proven correct.

However, since arguments can be false while the conclusion is right, the threat dualism poses to a scientific solution to the problem of phenomenal consciousness has not disappeared yet. What has been established in this chapter is that there is no route from the logical, epistemic, natural or metaphysical possibility to the physical possibility of zombies. So, at this moment, we should be neutral about their physical possibility, for only the argument in favor of zombies has shown to be false. In the next chapter I will – for the sake of argument – suppose that phenomenal zombies

---

2 As will become clear in the following chapters, I do not believe we are all that ignorant. I believe that zombies are physically impossible. For the sake of argument, I here accept that we are ignorant, because I want to argue against conceivability as an argument pro or contra any position that starts from scratch.
are physically possible, showing that the consequences of this assumption leads to a *reductio ad absurdum* not only is the argument in favor of zombies false, the notion itself is incoherent.
CHAPTER FOUR
DAVID CHALMERS’ PROPERTY DUALISM

In the preceding chapter we saw that if (tough) property dualism is right a scientific solution to the problem of phenomenal consciousness cannot be given, for science assumes monism. The main argument for this kind of dualism, the conceivability argument, has been shown to be invalid. However, the position itself might be right even though the argument in favor of it is not sound. In this chapter I will discuss the most recent version of property dualism – viz. David Chalmers’ version – and show that it leads to a reductio ad absurdum. Since the argument against his position is a complex one, I will start by presenting the relatively short version of it.

1. the short version of the argument
My main claim is that Chalmers’ theory is inconsistent for the following reason. His correct assumption that we know that we are phenomenally conscious contradicts with the consequence of his theory that we cannot know that we are phenomenally conscious. Since his initial assumption is correct, his theory must be false.

I will start by presenting Chalmers’ distinction between logical and natural supervenience. According to Chalmers phenomenal consciousness is merely naturally supervenient, which is natural supervenience without logical supervenience. This in contrast to the broadly physical properties that are all logically supervenient on the lowest physical subvenient base.

After the discussion of logical and natural supervenience, I will show that (1) there can be no examples of mere natural supervenience within the broadly physical domain and (2) that those examples that fall outside the broadly physical domain are not interesting for they cannot be known. This is not enough to show that Chalmers’ theory is incorrect, for he might argue that phenomenal consciousness is a very special merely naturally supervenient property, viz. one that can be known. This is an exception because we have a special epistemic relation to it; according to Chalmers we are directly acquainted with our phenomenal experiences (and not with other nonphysical properties). I will argue that this epistemic relation cannot save phenomenal consciousness from the earlier argument that we cannot know merely naturally supervenient properties, for (1) the relation is mysterious and introducing mysteries does not solve problems, and (2) the argument in favor for the acquaintance relation is a decoy: Chalmers tries to divert our attention to a related problem (that of the justification of phenomenal judgments) and then
presents a solution to that. I will argue that this maneuver takes us right back to the heart of the problem, showing that his decoy has exactly the same problem as his theory that phenomenal conscious is merely naturally supervenient.

As long as Chalmers does not solve this problem the argument that we cannot know any merely naturally supervenient property also applies to phenomenal consciousness: if a property is merely naturally supervenient, then we cannot know it. Since we do know we are phenomenally conscious, the conclusion has to be that phenomenal consciousness is not merely naturally supervenient, hence Chalmers' property dualism is incorrect.

2. **logical and natural supervenience**

2.1 **logical supervenience**

Chalmers characterizes logical supervenience as follows.

> When B-properties supervene logically on A-properties, we can say that the A-facts entail the B-facts, where one fact entails another if it is logically impossible for the first to hold without the second. (1996b: 36)

The use of the notion of logical impossibility in this characterization suggests that the notion of logical supervenience might be viewed as a conceptual rather than an ontological notion. In chapter two I distinguished two kinds of logical possibility: formal and conceptual possibility. Chalmers' notion of logical possibility belongs to the latter, since he says that “[i]n determining whether it is logically possible that some statement is true, the constraints are largely conceptual.” (1996b: 35) I also showed that this kind of possibility is not a guide to what is physically possible, and that one of the major problems with conceptual possibility is that it clearly depends on our conceptual scheme whether we consider a scenario to be conceptually possible or not.

Remember the giant insect example: at t = 0, the time we did not know that nowadays insects cannot become gigantic, giant insects were believed to be logically possible for there was no conceptual contradiction in this notion. Then we found out that nowadays giant insects are an ontologically impossibility, because of their physical make up and the present day atmosphere. Arguably, this new knowledge changes the concept ‘insect.’ If the concept does not change, then there are scenario’s that remain conceptual possible even though we know that the scenario is ontologically impossible. If this is the case, then conceptual possibility says nothing about ontological possibility.
If the new knowledge does change the concept ‘insect’ then giant insects become conceptually impossible. But we can only draw that conclusion after we gained the new knowledge about the insects. Again, conceptual possibility says nothing about ontological possibility.

So, in either case conceptual possibility, Chalmers’ preferred version of logical possibility, is clearly a conceptual notion and not an ontological one. He uses this notion in the definition of logical supervenience, suggesting that logical supervenience is a conceptual notion. However, there are two good reasons to interpret logical supervenience as a ontological notion.

First, Chalmers just is not very clear about what is conceptually possible or not. He explicitly allows for identification of logical with metaphysical possibility or possibility tout court (1996b: 35; 42). But metaphysical possibility – as I showed in chapter two – is a very different kind of possibility than logical possibility, for it evaluates whether a scenario is possible not by conceptual criteria, but by essential criteria. It is metaphysically impossible that gold is uncomposed, for gold essentially is composed. But being composed is a property of gold, independent of any concept we have of it. In chapter three I argued that these different evaluation criteria provide a good reason not to identify logical with metaphysical possibility. Here, I want to make another point. Since Chalmers allows the identification, his notion of logical supervenience becomes a metaphysical and therefore an ontological notion.

The second reason is that he also uses ontological terms in his characterization of logical supervenience, like “property” and “fact.” So, Chalmers clearly intends to be saying something about the real world, not just about our concepts of properties and facts. This is supported by the examples he presents to clarify the notion of logical supervenience.

At the global level, biological properties supervene logically on physical properties. Even God could not have created a world that was physically identical to ours but biologically distinct. There is simply no logical space for the biological facts to independently vary. When we fix all the physical facts about our world – including the facts about the distribution of every last particle across space and time – we will in effect also fix the macroscopic shape of all the objects in the world, the way they move and function, the way they physically interact. If there is a living kangaroo in this world, then any world that is physically identical to this world will contain a physically identical kangaroo, and that kangaroo will automatically be alive. (1996b: 35)

So, given the characterization and examples, it is evident that Chalmers’ notion of logical supervenience is an ontological notion that applies to what I call the broadly physical. A good example is the supervenience of shape. If we take eight cubes and stack them two by two by two, then we get another cube.
According to Chalmers’ terminology, logical supervenience and reductive explanation (Cl. chapter one) imply each other: if a property is logically supervenient it can be reductively explained, and if it is reductively explainable, then it has to be logically supervenient (1996b: 47-50). Since a property is reductively explainable if it can be analyzed functionally, which is an analysis in terms of the causal powers of the property (1996b: 44): logically supervenient properties have causal powers. This is not the case with merely naturally supervenient properties.

2.2 natural supervenience

Chalmers distinguishes logical from natural supervenience. Just as logical supervenience, natural supervenience also is an ontological notion. Chalmers starts by saying that “[t]here can be supervenience without logical supervenience” (1996b: 36). This is natural supervenience, which he characterizes as follows.

In general, B-properties supervene naturally on A-properties if any two naturally possible situations with the same A-properties have the same B-properties. (1996b: 36)

We can have both logical and natural supervenience at the same time. Given the definition of natural supervenience, we can have a situation that actually holds in our world. Chalmers’ position about phenomenal consciousness is that it does not supervene logically: it supervenes only naturally. “Where we have natural supervenience without logical supervenience, I will say that we have mere natural supervenience.” (1996b: 37)

Since Chalmers wants to show that phenomenal consciousness is such a merely naturally supervenient property, he cannot use it as an example to elucidate the notion. But we can use it to see what Chalmers wants to demonstrate. He uses different thought-experiments to support his claim that phenomenal consciousness is merely naturally supervenient, among which are the two scenario’s in which our counterparts in a physical duplicate of our world have altered (inverted) phenomenal consciousness (1996b: 99-101) or (2) have no phenomenal consciousness (1996b: 94-99). Let us call the worlds w₁, w₂, and w₃, respectively our world, the inverted world and the zombie world. All these worlds are microphysically identical duplicates of each other, which means that all the broadly physical properties are also the same, for they supervene logically in every world. In w₂, phenomenal experiences are inverted and in w₃ they are absent. So, merely naturally supervenient properties can be different or absent in another physically identical world.
B-properties supervene merely naturally on A-properties only if for any three worlds \( w_1, w_2 \) and \( w_3 \) identical with respect to their A-properties, it is possible that the B-properties that are present in \( w_j \) are different in \( w_2 \) or absent in world \( w_3 \). In other words, merely naturally supervenient B-properties can vary independently of their subvenient A-properties across worlds.

This characterization can be used as a test to evaluate whether a property \( X \) is merely naturally supervenient. If it is, it has to be a B-property that can vary across worlds, even though the properties on which it supervenes remain constant.

The reason that there is variety in phenomenal properties between the three worlds is that there are different sets of physical-phenomenal laws in these worlds. Chalmers calls these laws psychophysical laws, but those would be laws between the physical and the psychological. But according to Chalmers the psychological is logically supervenient on the physical, hence psychophysical laws cannot apply to phenomenal properties. In other words, Chalmers argues that the phenomenal falls outside what I call the broadly physical domain, while he claims that psychological features are broadly physical. Psychophysical laws evidently apply to the broadly physical domain. Hence, these laws cannot apply to the phenomenal. What he needs are (broadly) physical-phenomenal laws.

In our world there are physical properties, our physical laws and set # 1 of physical-phenomenal laws. In the inverted phenomenal consciousness world this latter set is displaced by set # 2 and in the zombie world such a set is absent. These different sets determine how the phenomenal is dependent on the physical world. This is where Chalmers differs from Descartes: according to Chalmers the phenomenal world (if it exists in a world \( w_j \)) is always dependent on the physical world, so ghosts – who are totally independent of matter – cannot exist. According to Descartes ghost can exist (1642: 91) and matter can exist without the res cogitans resulting in his substance dualism. Since Chalmers makes the phenomenal dependent on the physical, he clearly is not a substance dualist. He calls himself a property dualist, where the two sets of properties he is talking about are ontologically almost as different as Descartes’ two substances. The property dualism (or better pluralism) I defend is quite different from this property dualism, for mine is a variety of monism: all different properties belong to one domain – the broadly physical (Cl. chapter one). If I were to agree on Chalmers’ terminology I would claim that phenomenal properties are logically supervenient on the physical, just as biological or chemical properties are.
I will argue that this is what Chalmers also should claim if he takes his distinction seriously and remains a realist about phenomenal consciousness.

From Chalmers’ account of logical and natural supervenience, it follows that merely naturally supervenient properties cannot be reductively explained for ontological reasons: the same A-properties can be the subvenient base of different (or no) B-properties. This is what (T) expresses. Another way of putting it is this: since (T) states that similar A-properties can obtain across worlds, with or without certain B-properties that supervene on the A-properties, the B-properties can neither be reduced to nor be reductively explained by the A-properties; the B-properties are fundamental properties. Here we have properties that at first sight have a strange ontological status: they are both supervenient and fundamental. They are supervenient for they are dependent on broadly physical properties and they are fundamental because they cannot be reductively explained. The latter also implies that – contrary to fundamental properties that are not supervenient – they cannot have causal powers, for if supervenient properties had causal powers they could be reductively explained. So (T) can be extended in the following way:

\[ \text{(T)} \quad \text{B-properties supervene merely naturally on A-properties only if for any three worlds } w_1, w_2 \text{ and } w_3 \text{ identical with respect to their A-properties, it is possible that the B-properties that are present in } w_1 \text{ are different in } w_2 \text{ or absent in world } w_3. \text{ In other words, merely naturally supervenient B-properties can vary independently of their subvenient A-properties across worlds. These B-properties cannot be reductively explained, hence can have no causal powers.} \]

2.3 two kinds of supervenient properties

According to Chalmers there is no such thing as mere logical supervenience: logical supervenience implies natural supervenience. “If any two logically possible situations with the same A-properties have the same B-properties, then any two naturally possible situations will also.” (1996b: 37) This means that we seem to have but two options: either a supervenient property is logically and naturally supervenient, or the property is not logically but merely naturally supervenient. Things, however, are not this simple for Chalmers is not clear about the status of physical laws.

2.4 an ambiguity concerning physical laws
There is a problem concerning the status of physical laws. Chalmers is highly ambiguous about whether they are part of the physical facts or not, hence whether or not they are part of the subvenient base of higher-order properties.

It is also useful to stipulate that the world's physical facts include its basic physical laws. On some accounts, these laws are already determined by the totality of particular physical facts, but we cannot take this for granted. (1996b: 33)

The problem is that he sometimes includes them (as in the kangaroo example, Cf. supra), sometimes excludes them (as we will see in a moment) depending on what suits him. What are the consequences of either option for Chalmers’ theory?

If physical laws like the gas law do belong to the subvenient base, then the pressure of a mole of gas with volume \(x\) and temperature \(y\) in world \(w_i\) cannot be different from the pressure of a gas with the same volume and the same temperature in a duplicate world \(w_j\). On this account the pressure of a mole of gas is logically supervenient on its volume and temperature.

As we saw in section 2.1 Chalmers explains logical supervenience using the example of biological properties that supervene on physical properties: if the physical properties are fixed, so are the biological. A physical duplicate of a living kangaroo in a physical duplicate world will be a living kangaroo. This presupposes that the physical laws are included in the subvenient base. If they are not then the physical duplicate of the kangaroo might well be dead.

If physical laws like the gas law do not belong to the subvenient base, then the pressure of a mole of gas with volume \(x\) and temperature \(y\) in world \(w_i\) might be different from the pressure of a mole of gas with the same volume and the same temperature in a duplicate world \(w_j\). Since now it is logically possible for the A-facts to hold without the B-fact (the pressure in \(w_i\)), the gas law scenario cannot be an example of logical supervenience given Chalmers’ characterization of it. So, if the physical laws do not belong to the subvenient base, then the pressure of a mole of gas can vary independently of the facts on which it supervenes, hence it passes (T). On this account of physical laws, then, the pressure of a mole of gas is a merely naturally supervenient property. This is precisely the example Chalmers uses to explain this notion.

In the actual world, whenever there is a mole of gas at a given temperature and volume, its pressure will be determined: it is empirically impossible that two distinct moles of gas could have the same temperature and volume, but different pressure. It follows that the pressure of a mole of gas supervenes on its temperature and volume in a certain sense. (In this example I am taking the class of A-properties to be much narrower than the class of physical properties, for reasons that will
become clear.) But this supervenience is weaker than logical supervenience. It is *logically* possible that a mole of gas with a given temperature and volume might have a different pressure; imagine a world in which the gas constant $K$ is larger or smaller, for example. Rather, it is just a fact of *nature* that there is this correlation.

This is an example of *natural* supervenience of one property on others: in this instance, pressure supervenes naturally on temperature, volume, and the property of being a mole of gas."

(1996b: 36)

In sum, when Chalmers presents us the kangaroo example to explain the notion of logical supervenience, the physical laws are included in the physical facts, when he presents the gas law example to explain the notion of natural supervenience the physical laws are excluded. Chalmers, then, is ambiguous as to whether or not the physical laws are included in the physical facts. This presents him with the following dilemma.

Chalmers’ first option would be to claim that the physical laws do not belong to the physical facts. This would imply that the kangaroo example is wrong, for a physical duplicate of a kangaroo might not be alive, just as a physical duplicate of a mole of gas with temperature $x$ and volume $y$, might not have the pressure it has in the world that is a physical duplicate our world. This would make all broadly physical properties merely naturally supervenient.

Chalmers’ second choice is to claim that the physical laws do belong to the physical facts. This would imply that the gas law example is wrong, for a physical duplicate of a mole of gas under the same circumstances would have to have the same pressure, for duplication of the physical facts would also duplicate the physical laws.

Since physical laws are either physical facts or not, Chalmers has to make a choice. I believe that he would choose that the physical facts include the physical laws. There are two reasons for this. The first reason is that it only conflicts with the gas law example and does not give up the usefulness of the distinction between logical and natural supervenience as ontological notions. Choosing for the first option would precisely give up this distinction, for in that case all supervenient properties are merely naturally supervenient, leaving no room for logical supervenience as an ontological notion. Everyone then would readily agree that phenomenal consciousness is a natural supervenient property. (Of course logical supervenience as a conceptual notion would still hold for those cases that belong to the domain of mathematics and logic, but that is not the way these notions are used: Chalmers chooses his examples of logical and natural supervenience in the domains of the physical, chemical, biological, psychological and phenomenal, not in the domains of the mathematical and logical.)
2.5 there are no examples of mere natural supervenience in the broadly physical domain

The second reason is that including the physical laws in the physical facts is coherent with Chalmers’ views on reductive explanation, while excluding them is not. According to Chalmers only logical supervenient properties can be reductively explained (1996b: 48). Including the physical laws in the subvenient base is necessary for a reductive explanation of – say – pressure, for if one does not, one cannot explain how a certain pressure is realized by its subvenient base, for the same subvenient base can exist without the supervenient property. This comes down to saying that in the broadly physical domain there cannot be any examples of mere natural supervenience, a claim Chalmers seems to be sympathetic to for he says that: “it is hard to find cases of natural supervenience on the set of physical properties without logical supervenience.” (1996b: 37) Furthermore, he suggest that: “[m]odulo conscious experience, all phenomena are logically supervenient on the physical.” (1996b: 76)

In sum, I have argued that Chalmers’ notion of logical supervenience is not a conceptual but an ontological notion, and that physical laws should be included in the subvenient physical facts, for otherwise every supervenient property would become naturally supervenient. This is supported by Chalmers’ views on reductive explanation and his remarks on finding examples of mere natural supervenience. Given this, examples of mere natural supervenience can only be found outside the domain of the broadly physical. In the next section I will show that such examples are not intelligible.

2.6 there are no intelligible examples of merely naturally supervenient properties outside the broadly physical domain

The fact that there are no examples of mere natural supervenience in the broadly physical domain does not mean that there are no examples of such properties at all. There might be examples outside this domain. To be sure, that is exactly what Chalmers wants to prove: phenomenal consciousness is a property that is both (1) nonphysical and (2) merely naturally supervenient on the physical. Since Chalmers would like to prove this, he is not in a position to use phenomenal consciousness as an example of how a property can be merely naturally supervenient. What other kind of properties might be used as examples of this notion? I will argue that there are no other examples, for we cannot know them; epistemic access to a property requires a causal relation between the knower and the property known. Merely naturally supervenient properties cannot be known, for they have no causal powers. This means that in the following section I will have to discuss whether it makes sense that phenomenal consciousness is the only merely naturally
supervenient property that is epistemically accessible, via something else than a causal connection. I will evaluate Chalmers answer, that this epistemic relation is acquaintance.

Let us suppose – for the sake of argument – that there are ectoplasmatic entities that supervene merely naturally on some physical property, say pineal glands, in a world \( w_2 \) physically identical to our world \( w_1 \). Furthermore, suppose that ectoplasmatic entities do not supervene in our world, and that in both worlds humans are phenomenally conscious. To pass (T) ectoplasmatic entities have to be fundamental properties in the sense described above: they are dependent on the physical for their existence, but can neither be reduced to nor reductively explained by the physical. What is important here, is that they cannot have any causal influence on the world in which they supervene; if a supervenient property has any causal influence on its subvenient base, then it has to be a logically supervenient property, for then a reductive explanation is possible (Cf. supra).

This demand has its implications for a theory of knowledge. I believe that a causal theory of knowledge is correct: to have knowledge about any property there has to be a causal chain between the property known and the person who knows that property. A causal theory of knowledge together with the demand – in order to pass (T) – that the ontologically real ectoplasmatic entities of world \( w_2 \) are causally inefficacious, implies that the humans in the world with these entities cannot know that ectoplasmatic entities exist or that they supervene merely naturally on some property in their world. So, there has to be another way of gaining knowledge if Chalmers wants to stick to his claim that phenomenal consciousness is a merely naturally supervenient property that is known.

Chalmers indeed explicitly rejects the causal theory for things that we do know for certain (1996b: 193-196) and claims that we are immediate acquainted with our phenomenal experiences (1996b: 196-198). This direct acquaintance is a noncausal relation between thing known and the knower. Let us for the sake of argument accept that we are able to have knowledge by means of an immediate, noncausal, acquaintance relation. Is accepting acquaintance a way to show that phenomenal consciousness passes (T)? I will argue that it is not.

In Chalmers’ theory judgments and knowledge states belong to the domain of psychological consciousness (1996b: 26). Psychological consciousness is causally efficacious, logically supervenient and reductively explainable. This means that a change in knowledge state cannot
occur without a change in subvenient base (if it would, that would be passing (T) and the knowledge state would be merely naturally supervenient).

Gaining knowledge about a merely naturally supervenient property evidently is a change in knowledge state, hence a change in subvenient base. What caused this change? It cannot be a phenomenal experience, since all phenomenal experiences are causally inefficacious. The cause had to be something else, something physical. This means that in any physical duplicate of the world that contains the merely naturally supervenient property, this change would also occur, even if the merely naturally supervenient property is not present. This would mean that if our physical duplicates of \( w_2 \) claim to have knowledge of merely naturally supervenient properties like ectoplasmatic entities, we should claim this too. However, humans do not claim to have knowledge of any merely naturally supervenient properties, with the exception of phenomenal consciousness. This implies for all alleged naturally supervenient properties other than phenomenal consciousness either that:

1. they do not exist at all. This would indeed explain that we are not claiming to know that ectoplasmatic entities exist. This would render the notion useless.

Or

2. they do exist but cannot be known at all. This also renders the notion useless, for we will never be able to pick out a property in our world to apply the notion to.

2.7 is there a knowable merely naturally supervenient property after all?
The only option for Chalmers left is to claim that the only nonphysical merely naturally supervenient property that can be known is phenomenal consciousness. The way we have epistemic access to phenomenal consciousness has to be a special – noncausal – way. Chalmers indeed claims that this epistemic relation is such a special relation, which he calls acquaintance. This way of looking at it, is the only way in which the notion of mere natural supervenience can be useful. So far, there is no real problem for Chalmers, for he only needs to show that phenomenal consciousness is merely naturally supervenient. The fact that there can be no other examples does not really matter: he even seems to agree that phenomenal consciousness is the only merely naturally supervenient property \( \text{cf. supra} \). However, this leads to the problem of phenomenal judgment.
2.8 the problem of phenomenal judgment

The problem of phenomenal judgment is the following. We just saw that the alleged merely naturally supervenient properties are causally irrelevant to the judgments and knowledge states about them. Since the only candidate for a merely naturally supervenient property left is phenomenal consciousness, this is only a problem for knowledge states and judgments about phenomenal consciousness. “It seems that consciousness is explanatory irrelevant to our claims and judgments about consciousness. This result I call the paradox of phenomenal judgment.” (1996b: 177)

Chalmers frames the problem in terms of explanation. But the idea of course is that since phenomenal consciousness is merely naturally supervenient it cannot be reductively explained for ontological reasons: it has no causal efficacy, and reductive explanations are only possible for properties that do have causal efficacy. The problem with both causal and explanatory irrelevancy of phenomenal consciousness for phenomenal judgments is the question: how can they qualify as knowledge, for zombies make the same phenomenal judgments, but we would not qualify these as knowledge? (1996b: 192)

2.9 the problem of justification of phenomenal judgments

Chalmers explicitly puts the problem of the causal irrelevancy of phenomenal consciousness aside (1996b: 177) and tries to show how phenomenal experiences are relevant to phenomenal judgments: they justify them. In this way Chalmers tries to save phenomenal judgment from not qualifying as knowledge in our case. In the case of the zombie of course, their judgments are not justified: “But surely his [the zombie’s] judgments are not justified at all. After all, they are utterly and systematically false.” (1996b: 192)

There are two problems with this problem of justification of phenomenal beliefs. The first is that it functions as a decoy: it tries to divert our attention from the real problem, which is how we can have knowledge of phenomenal experiences if they do not have any causal efficacy, while a knowledge state belongs to the broadly physical domain, hence has to have a broadly physical cause. The real problem for Chalmers then is to show how acquaintance can solve the problem of phenomenal judgment. As long as he does not do that, we have no reason to believe that phenomenal consciousness is any different from other alleged merely naturally supervenient properties: if it exist it cannot be known. Since it exists and we do know it, the theory has to be
false and will not be saved by invoking a mysterious relationship: introducing a mystery does not turn a contradiction into a paradox.

The second problem is that Chalmers cannot avoid the problem, for he tries to push objective criteria into subjective epistemology, which is both a fallacy and leads him straight back to the problem of phenomenal judgment. I will first show that his account of the justification of phenomenal judgments is an objective account, which he then presents as subjective. After that it will also be clear that this account leads back to the problem he tried to bypass earlier.

### 2.10 the fallacy and backfire

What justifies our phenomenal judgments? Chalmers says it is *having* the experiences that justifies the beliefs. “[T]he fact that I have a red experience now provides justification for my belief that I am having a red experience.” (1996b: 196) He calls this relation “acquaintance” and claims that it is an intimate epistemic relation to the experience (1996b: 197).

I agree that – seen from an *objective* point of view – the *having* of conscious experiences makes the phenomenal judgments true. But justification of one’s own judgments is not an objective matter. Let me give some examples. It is the unicorn in Africa that makes my claim that there is a unicorn in Africa true. It is the existence of god that makes Descartes’ claim that god exists true. These are truth conditions. But in order to use the truth of one’s claims to justify them, *subjective* knowledge that the truth conditions are satisfied is needed. The problem, then, is to know the facts that can be used to justify the claims. For me to know that there is a unicorn it is not enough that there is a unicorn, but I must also have an epistemic relation to the unicorn. But I have never seen a unicorn, nor do I stand in any other epistemic relation to unicorns. So I cannot know that my claim is true; I can only state the truth conditions and I do not know whether these conditions are satisfied. Similarly, when Chalmers says that it is the having of conscious experiences that justifies his phenomenal judgments he is stating the truth conditions. He needs to show that he also knows that the truth conditions are satisfied and how he knows this, just as I would have to do if I were to claim that I know that unicorns exist.

The implication of the difference between knowing what the truth conditions are and knowing that they are satisfied, is that Chalmers does not solve any problem by saying that:
To note that in a different case the belief might have been formed in the absence of the evidence is not to say that the evidence does not justify the belief in this case. I know that I am conscious, and the knowledge is based solely on my immediate experience. (1996b: 198)

At first sight Chalmers seems to get things right here: to justify one’s judgments, one at least needs to know that the truth conditions are satisfied. However, this is not the case, for the judgment that needs to be justified is “I know I am phenomenally conscious” which can only be justified by explaining how one knows that the truth conditions are satisfied. Just stating that one knows this is a trick. What happens here is that Chalmers tries to justify his phenomenal judgment “I am phenomenally consciousness” with the phenomenal judgment “I am phenomenally consciousness.”

Here, the problem of phenomenal judgment comes back to haunt him, for a zombie would make exactly the same claim. In order for this phenomenal judgment to be true, knowledge that its truth conditions are satisfied is needed. This means we have to know that we are conscious. Now we are back were we started: how do we know we are conscious? Chalmers’ alleged solution, then, is question begging: The question how we know we are phenomenally conscious, cannot be solved by answering that we have knowledge that we are phenomenally conscious. What is needed is a specification of how we gain such subjective knowledge. According to Chalmers the answer is that we know this by acquaintance. He does not even begin to specify this notion. This is not acceptable: the relation should be specified. I do not see how this can be done. If the paradox cannot be solved, then the conclusion has to be that it is not a paradox but a contradiction, hence that Chalmers’ theory just cannot be right.

The problem with Chalmers’ solution is not the same as the iteration problem. The iteration problem is that if one asks how someone knows something, he has to answer the question how he knows that, etc. If the problem of the return of the paradox of phenomenal judgment were this problem, it would be very unfair to challenge Chalmers on this point. Chalmers’ problem is worse. He has no explanation how we might have knowledge of a naturally supervenient property in the first place. A theory of consciousness that accepts a causal relation between properties and the intentional states about these properties has at least a specified relation that is able to solve this problem in principle. Chalmers has to specify the acquaintance relation, for it is presented as a relation that is constitutive for knowledge of phenomenal states. At this moment Chalmers has not done so. We all should just believe that an utterly mysterious epistemic relation called “acquaintance” will solve the problem of phenomenal consciousness.
2.11 leaving the problem of phenomenal judgment unsolved

However, let us again grant Chalmers all the benefit of the doubt, and keep the A-facts of both our world and the zombie world identical. This leaves the problem of phenomenal judgment unsolved; it is a challenge Chalmers has to take up at some other time. We know that we are conscious, and zombies do not know that. This leads to an even weirder scenario. The zombie claims to know that he is conscious, but *ex hypothesi* he is not. Since – according to Chalmers – a knowledge state is a physical state, zombies must find out that they are not conscious, in the very same way that they (and we) find out they have no knowledge of ectoplasmatic entities or any other naturally supervenient property.

Since having knowledge is a psychological state, it can be the cause of an action, like uttering: “Hmm! Something strange just happened! I claim to know that I am conscious, that I have direct acquaintance with my conscious experiences. What do I mean by that?” I do not see any reason to keep the zombies from refraining to asking this question. In answering this question they will pretty soon find out that their knowledge claim was false: they do not know that they are conscious. Both this knowledge state and the utterance change the A-facts of the zombie world. To keep conscious experiences naturally supervenient, *our world has to change accordingly*. But we never find ourselves baffled like the zombies do that we claim to have knowledge of conscious experiences that we know we do not have. This leads to just one conclusion: the zombie world is not A-fact-identical to our world, hence consciousness cannot be merely naturally supervenient, for the subvenient base changes with the supervenient properties: *phenomenal consciousness fails* ($T$).

2.12 there are no knowable merely naturally supervenient properties

So, the alleged possibility of knowledge of a particular merely naturally supervenient property shows that the property cannot be merely naturally supervenient. This means that the last property that might have been known to be so cannot in fact be known to be merely naturally supervenient. This leads to the following conclusion. If we have knowledge of a supervenient property, it has to be a logically supervenient property. This conclusion has two aspects. The first is an *epistemic* one: since we will never be able to *know* that a particular property is merely naturally supervenient, the notion of merely natural supervenience is useless, for it would only be useful if we were able to attribute it to particular properties. The second part of the conclusion is *ontological*: any property that is known has to be either logically supervenient or fundamental.³

³ This raises the question whether these fundamental properties might be naturally supervenient. The answer is no. First, all naturally supervenient properties have to be fundamental by definition, but this does not mean that fundamental properties are naturally supervenient, for I take it that most fundamental properties are just

61
Known merely naturally supervenient properties do not exist. Since phenomenal consciousness is a property that is known, it cannot be merely naturally supervenient. The further implication is that it has to be logically supervenient, which in turn implies that it has to be reductively explainable, and that it is not a fundamental property of our world.

We are now in a position to say that the most frequently used argument for dualism is false and that the most recent version of tough property dualism is incoherent, which provides us with good reason not to believe in the truth of dualism, even though most people live in cultures that do so. Tough property dualism is no longer a threat to a scientific solution of the problem of phenomenal consciousness. This does not mean that we now can go on and try to solve the problem, for there are also arguments thought up by agnostic and monist views that purport to show that a scientific solution is not possible. I will discuss the agnostic argument in chapter five and the monist argument in chapter six.
CHAPTER FIVE
NO HARSH DEMANDS

In this chapter I will address the related arguments of Thomas Nagel (1974) and Frank Jackson (1982; 1986) that both purport to show that a physicalist theory leaves phenomenal consciousness out, whether it is at present, as Nagel argues, or permanently, as Jackson claims. Evidently these arguments need to be addressed for if physicalism indeed leaves phenomenal experiences out, we should either opt for Jackson’s view that a scientific solution cannot be given ever, or for Nagel’s view, that we should be agnostic about the possibility of ever solving the problem, and that we cannot do this right now.

Nagel’s and Jackson’s arguments are related for they both demand that a theory of consciousness should tell us what it is like to have phenomenal experiences such as the experiences of red or of what it is like to be a bat. John Locke wrote in his An Essay Concerning Human Understanding (1690) that

A studious blind man, who had mightily beat his head about visible objects, and made use of the explication of his books and friends to understand those names of light and colours which often came in his way, bragged one day, that he now understood what scarlet signified. Upon which his friend demanding, what scarlet was? The blind man answered, It was like the sound of a trumpet. (Book III, Chapter IV, Section 11)

If a theory about consciousness a theory of phenomenal consciousness should tell us what certain experiences are like, the theory should be able to tell the blind man what it is to see scarlet without invoking sensations of other sensory modules. However, on this point I side with Gerald Edelman and Giulio Tononi (2000) that this demand is just too strong.

What is the problem that Thomas Nagel (1974) and Frank Jackson (1982) bring to our attention? Basically they argue that a physical theory leaves the most important thing out, viz. what is it like to have an experience of a certain kind, or more general what it is like to be an experiencing subject. Not everyone agrees that this should be or even can be the goal of a theory of consciousness.

No amount of description will ever be able to account fully for a subjective experiences, no matter how accurate that description may be. Many philosophers have used the example of color to make their point. No scientific description of the neural mechanisms of color discrimination, even if it is perfectly satisfactory, will make you understand what it feels like to perceive a particular color. (Edelman & Tononi 2000: 11)
One might feel that evidently the demand is too strong, and that neither Nagel nor Jackson really
demand that a theory should give us the knowledge what it is like to be a bat, to see color or to
hear a trumpet. Yet, the following quote from Nagel really makes this strong demand:

Though presumably it would not capture everything, [the goal of an objective phenomenology]
would be to describe, at least in part, the subjective character of experience in a form
comprehensible to beings incapable of having those experiences. (1974: 449)

Earlier in the same article this strong demand is also expressed by Nagel when he says “I want to
know what it is like for a bat to be a bat. (1974: 439)” Since no description can ever meet this
demand, Nagel’s conclusion is that we cannot understand how physicalism can be true, even if it is
ture: we lack the right concepts to link those that refer to the subjective to those that refer to the
objective (1974: 446).

**Missing every shade of blue**

Once, Jackson went even further. He claimed that based on the fact that a physicalist theory
cannot meet the demand physicalism has to be false (1986: 295). This argument is know as the
knowledge argument and makes use of the following scenario. Consider Mary, a brilliant scientist
grew up in a black and white room, which we might call the Jackson-room (Perry 2001: 19).
Leaving aside practical issues of how to hide the color of her skin and prevent her from bleeding
when she cuts herself, this thought-experiment assumes that she has never had an experience of
any color besides white, black and shades of gray. Furthermore, it is supposed that we have a
complete physical theory of colors, and that Mary knows this theory. So, Mary knows every
physical fact there is to know about colors, but she still does not know what it is like to experience
blue. She gains this piece of knowledge when one sunny day she leaves the Jackson-room and
looks at the sky. Since she already knew all the *physical facts* about colors, the new knowledge what
it is like to experience blue cannot be knowledge of a physical fact about colors, hence even a
complete physical theory of colors leaves something out, viz. the phenomenal feel. Hence,
physicalism stands refuted (Jackson 1982, 1986; Chalmers 1996b; Gertler 2000). My view on
Jackson’s knowledge argument is quite simple: the premise that Mary has *all* the knowledge about

---

4 Although Nagel has made this phrase famous, Brian Farrell actually used the same phrase to point out the essential
problem of consciousness: “I wonder what it would be like to be, and hear like, a bat.” (1950: 183)
5 In his *From Metaphysics to Ethics* (1988) Jackson has given up this line of argument, as too speculative: we should not
give our “intuitions about possibilities too big a place in determining what the world is like.” (1998: 43-44). This is
why Jackson now has doubts about the knowledge argument.
the physical facts is simply wrong: she lacks some of the knowledge of some fact that concerns the broadly physical domain. Let us reconstruct the argument schematically:

1. If physicalism is right, there are no facts other than the physical facts;
2. Mary knows all the physical facts about colors;
3. When leaving the Jackson-room Mary gains knowledge about facts concerning colors;
4. The additional facts are nonphysical facts about colors [from (2) and (3)];
5. Physicalism is false [from (1) and (4)].

As we have seen Jackson accepts that a physicalist may include in the ontology all the properties that supervene on the fundamental level (Cf. chapter one). Hence, we can interpret this as follows: physicalism entails the claim that all *broadly* physical facts can be objectively described and then learned by scientist like Mary. It is here that the problem lies. Let me explain by adding someone to the scenario. Suppose Mary had a twin-sister, Mary-Ann, just as brilliant as Mary, who grew up in an environment where she was allowed to experience colors. Mary-Ann had the same interest in colors, and has read everything about colors. There is not a thing about colors Mary knows that Mary-Ann does not know (but not vice versa). After having studied every book about colors, they gain interest in some philosophical problems about them. They discuss Hume’s problem of the missing shade of blue. Of course Mary-Ann knows what it is like to experience Hume’s missing shade of blue. Mary-Ann wants her sister to know what that is like and writes her a letter (in black ink on a white sheet of paper), describing the missing shade of blue, concluding that: “*this* is Hume’s missing shade of blue.” As we all know, this can’t be done, and Mary will still miss every shade of blue. Demanding that a physical theory can do exactly this is absurd. It is analogous to demanding from a theory about liquidity not only that it describes what it is and when it occurs, but also that it actually constitutes liquidity. So, I agree wholeheartedly with Edelman and Tononi as they say that

> Scientific explanations provide the conditions that are necessary and sufficient for a phenomenon to take place, can explain the phenomenon’s properties, and can even explain why the phenomenon only takes place under those conditions. But no scientific description or explanation can substitute for the real thing. (2000: 12)

Trying to explain what it is to experience red to a congenitally blind man, therefore, will never work: “objective phenomenology” (Nagel 1974: 449) is a contradiction in terms. Hence Nagel is wrong when he says that the development of an objective phenomenology should not start by trying to make humans understand bat-experiences, but by trying to explain to a person blind
from birth what it is like to see. He is right however when he says that the analogy “Red is like the sound of a trumpet” is of little use. “That should be clear to anyone who has both heard a trumpet and seen red.” (1974: 449)

In sum, both the argument of Nagel and that of Jackson make the demand that a physicalist theory should tell us what it is like to be an experiencing subject. Since this demand is too strong the consequence that (we cannot understand how) a physicalist theory cannot be right does not follow.

Even though the demand that a theory about phenomenal experience should tell Mary or a blind man what it is like to see red is too strong, what should we say about real facts that cannot be described within a broadly physicalist framework? I believe that the solution lies in the rejection of premise two and that we should say that Mary does not know all the facts. She does not know what it is like to experience red, which is a (broadly) physical fact. This does not imply that a physicalist science cannot study these facts: it just means that they cannot be described in a manner that they are constituted in someone’s head by reading the description. That they can be studied within a broadly physicalist framework will be shown in chapters six through eleven.

To put it another way: it is a mistake to think that all broadly physical facts are objective. Surely, it is an objective fact that you have phenomenal experiences, but it is also an objective fact that these experiences have a subjective feel. How this feels like is something that can never be described in objective terms for that would constitute the phenomenal experience, but that is not what a theory should do: it should explain in objective terms why you have subjective experiences. That is something that is – in principle – not a problem for a broadly physicalist theory.

Having refuted the main argument for tough dualism, dualism itself and the famous arguments of Nagel and Jackson there is just one more argument that needs to be refuted: the argument from Colin McGinn, who agrees that there is nothing mysterious about consciousness, that the problem of consciousness can in principle be solved, but that we are just not smart enough to do so. I will turn to this argument in the next chapter.
CHAPTER SIX
THE NATURAL METHOD

In the preceding chapters I argued against dualist and agnostic positions that claim that there is no physicalist solution to the problem of phenomenal consciousness. But even within a monistic, naturalistic framework there are problems as Colin McGinn (1991) points out, for we might be cognitively closed for certain properties and theories. In this chapter I will briefly present McGinn’s reason for thinking that the scientific problem of consciousness cannot be solved. After this I will present Owen Flanagan’s natural method that is both an explicit answer to McGinn’s argument and the way I believe we might – step by step – solve the problem of phenomenal consciousness.

1. anti-constructive naturalism

McGinn is a noumenal naturalist. That is, he believes that consciousness is a natural phenomenon that could – in principle – be explained by a naturalistic, scientific theory. But, so McGinn argues, our cognitive capacities are too limited to understand this theory, just as a monkey’s cognitive capacities are insufficient to understand math, physics, or biology. The theory that describes how “technicolor phenomenology [can] arise from soggy gray matter” (1991: 1) is cognitively inaccessible for humans but might be accessible to other beings with greater cognitive capacities. Even though he is a naturalist and realist about consciousness, he argues that we cannot construct a theory that explains it. This is McGinn’s anti-constructive naturalism. How does McGinn defend this position?

McGinn describes cognitive closure as follows.

“A type of mind $M$ is cognitively closed with respect to a property $P$ (or theory $T$) if and only if the concept-forming procedures at $M$’s disposal cannot extend to a grasp of $P$ (or an understanding of $T$)” (1991: 3)

For instance, a monkey’s mind will be closed to the property of being an electron. The same goes for humans and the mind-body relation: we cannot know the property $P$ (nor theory $T$ about it) instantiated by the brain in virtue of which the brain is the basis of consciousness.

McGinn argues that there are two ways of identifying $P$: (1) via introspection or (2) via neuroscience. Neither of them will lead to $P$. First of all, introspection does not reveal that consciousness is
dependent upon anything physical: "Pure phenomenology will never provide the solution to the mind-body problem" (1991: 8). I have no problem with this claim.6

The claim that neuroscience cannot get us to \( P \), starts with the fact that consciousness is not an observable property of a brain. McGinn also argues that \( P \) also is an unobservable property of the brain. "I hereby invite you to try to conceive of a perceptual property of the brain that might allay the feeling of mystery that attends our contemplation of the mind-brain link." (1991: 11, my italics) McGinn thinks we cannot do this, for we are equipped with the capacity to represent the spatial world, and spatial properties are the kind of properties that cannot do the job of explaining how consciousness arises from the brain. Therefore, he concludes that neither consciousness itself, nor \( P \) can be perceived. But does this perceptual closure also lead to the cognitive closure of consciousness and \( P \)?

It is not true that generally perceptual closure implies cognitive closure, as McGinn rightly notices, for we cannot observe electrons, but can understand what they are. Using inference to the best explanation, we can come to know things that are unobservable, as long as they belong to the same kinds of things that we can observe. So, why should perceptual closure with respect to consciousness or \( P \) entail cognitive closure? The answer is: because consciousness is not a physical property, and it was the physical properties we were designed (by evolution through natural selection) to understand. Since the domain of the physical is causally closed, we will never need to introduce the concept of consciousness to explain anything we observe in this domain, for everything physical has a physical explanation. This is the homogeneity constraint on concept introduction (1991: 12-13). Hence, McGinn concludes, we are cognitively closed with respect to how we could introduce the concept of consciousness in the physical domain. The same goes for property \( P \): either \( P \) is a physical property and we cannot see how it is responsible for consciousness, or we introduce \( P \) as a nonphysical property, in which case we might as well have introduced consciousness itself.

The main idea is that the mental and the physical are described by two different registers of concepts. McGinn is in fact describing conscious inessentialism (Cf. chapter one): everything in the physical domain seems to be explainable without an appeal to consciousness. Hence, there is no need to introduce concepts that refer to it into a theory about human action.

---

6 "Premise (1) seems to me uncontroversially right. Consciousness reveals nothing about my brain, not even that I
2. **constructive naturalism**

There are several ways to argue against McGinn. I will present three. First of all, McGinn uses a *conceivability-strategy* to convince us of his view. He asks us to try to *conceive* of a property of the brain that would remove the feeling that the mind-brain link is mysterious. This is strategy, as I showed in chapter three, does not lead us away from a mystery or an ignorant position: The fact that we are ignorant *now* does not imply that we will be ignorant *forever*. This point can be stressed by using two analogies that have been used to describe our ignorant situation concerning the mind-body theory. As just noted, McGinn compares us with monkeys trying to understand physics: they will never be able to do that. Thomas Nagel, however, seems a bit more optimistic and compares us to the ancient Greek who is told that matter is energy (1974: 447). I prefer this optimistic view.

The second argument comes from Owen Flanagan. He argues that McGinn’s homogeneity constraint is overly restrictive. How can McGinn claim that we may not *introduce* the concept of consciousness, when it *has already been introduced*? The mind-body problem precisely exists because both facts about brains and facts about consciousness are on the table. We want to explain these facts and their relation. If we would accept the homogeneity constraint, then this “would lead us back into the dark ages of science.” (1991: 340) Flanagan argues that we need reports of conscious experiences to see what has gone wrong in – say – blindsight cases. Blindsight patients will claim not to see anything in a certain spatial area, but when forced to guess, their guesses show that they have preserved some of the capacity to pick up visual information of that area. The fact that they have lesions in area V1 of the visual cortex is a neurological fact that is used to explain why the conscious experience of vision is missing. So, it is strange to claim we should not use the concepts referring to conscious states when we are searching for scientific answers to consciousness problems.

The third argument is also from Flanagan. It starts with a counterclaim to McGinn: McGinn sees only two options – introspection or neuroscience – to solve the mind-body problem, thereby overlooking the method that might lead to a scientific solution: the *natural method*, which combines different methods. To show that McGinn is wrong indeed, Flanagan has to apply the natural method, solving the problem of consciousness bit by bit. In *Consciousness Reconsidered* Flanagan writes:

> have one!” (Flanagan 1990: 337)
This book is an attempt to make less puzzling the idea that consciousness is a natural phenomenon. I present a view of consciousness that I call "constructive naturalism." Consciousness is a natural phenomenon, and we can construct a theory about its nature, forms, roles, and origins by blending insights from phenomenology, psychology, cognitive science, neuroscience, and evolutionary biology. Consciousness is neither miraculous nor terminally mysterious. (1992: xi)

So, all of Flanagan’s work in which he successfully develops his constructive naturalism can be taken as an argument to show that McGinn is wrong. As I said in chapter one, I expect that the solution to the problem of phenomenal consciousness is not one that can be written down in one book. I believe that much of the data about the brain, our psychological life, our behavior etcetera have to be gathered and we need to evaluate time and again what the relevance is to the problem of phenomenal consciousness. Flanagan’s natural method is an excellent way to evaluate these data.

3. the natural method

For the last decade Owen Flanagan has developed and used the natural method (e.g. 1992, 1995a, 1995b, 2000). I am convinced that if we can solve the problem of phenomenal consciousness this method will do the job. The method is a multidisciplinary one: we should use phenomenology, cognitive science, and neurology to answer questions about consciousness. Additionally we should also make use of philosophy, evolutionary biology, psychiatry, anthropology and sociology if these disciplines provide answers to specific questions. I will present the natural method using one of Flanagan’s own examples: the splitting auditory attention.

On different occasions Flanagan describes the natural method as follows.

Start by treating three different lines of analysis with equal respect. Give phenomenology its due. Listen carefully to what individuals have to say about how things seem. Also, let the psychologist and cognitive scientists have their say. Listen carefully to their description about how mental life works and what jobs consciousness has, if any, in its overall economy. Finally, listen carefully to what the neurologists say about how conscious mental events of different sorts are realized, and examine the fit between their stories and the phenomenal and psychological stories. (1992: 11; 1995a: 1104; 1995b: 7-8)

4. splitting auditory attention

Flanagan uses the following experiment to show how the natural method works. Subjects are asked to pay attention to the left channel of a headphone. In this channel a sentence is presented, for instance: “the lieutenant put out the lantern to signal the attack.” This is a ambiguous sentence, for it might be interpreted either as (1) the lieutenant extinguished the flame of the lantern, or as (2) the lieutenant put the lantern outdoors.
The phenomenology of splitting auditory attention

When there is only noise presented in the right channel of the headphone, subjects do not show a particular preference for any of these interpretations. However, if among the noise in the right channel the sentence “he extinguished the flame” is also presented, a preference for interpretation (1) is shown, even though subjects still claim not to have heard anything meaningful in the right channel.

Taking the phenomenology seriously, we need to explain how this report can be reconciled with our knowledge that we did vary only one condition – the input of the right channel – that clearly resulted in the preference for interpretation (1).

The psychology of splitting auditory attention

A psychologist might come up with the following explanation. The sentence in the right channel never was conscious, which explains why subjects report that they have no memory of hearing it. The fact that the right channel sentence was relevant, then, has to be explained via unconscious processes that influence the processing of the target sentence.

An alternative psychological explanation might be that the right channel sentence was conscious, but not long enough to be remembered. This would also explain that subjects have no recollection of the sentence ever being conscious: they did not store that information. However, their recollection would be wrong. In this explanation the information presented at the right channel was consciously processed.

Here we have two psychological hypothesis that exclude each other and that are both coherent with the phenomenal data. These data then cannot help us in deciding which hypothesis (if any) is right. Hence, we need to gather more data.

The neurology of splitting auditory attention

How can we find out which one of these theories is right or that they both are wrong? Here neurology might help us. Crick and Koch (1990) have suggested that the groups of neurons that realize certain conscious states have synchronized 40 Hz firing patterns. I will discuss this suggestion in more detail in chapter nine. Suppose for the moment that this suggestion turns out to be true, then we can see how neurology plays an important role in this particular consciousness problem. Say that neurologists find the synchronized 40 Hz firing patterns for a
very brief period in the brain area that is known to be necessary for processing any auditory stimuli coming from the right ear. In that case we would have evidence in support of the second psychological explanation of the results of the experiment.

This merely is an example of how the natural method might work: given the phenomenology and the two psychological hypotheses, we know what the neurological evidence for either of the two psychological hypotheses might look like. Whether or not the second psychological hypothesis indeed turns out to be true, an important point is made by what is suggests. The point is that we need to add different disciplines to phenomenology, for we might well be wrong about parts of our own phenomenal life. In this case we would be wrong about our memories of our past phenomenal experiences. We can also be wrong about present phenomenal experiences. If one has a pain, then this experience is immune to error: we cannot be wrong that we experience what we experience. However, we can be wrong about the source of the pain:

Some pain is ‘referred pain.’ The source of what seems to be lower back pain is sometimes not strained or knotted muscles in one’s back, but a growth or misalignment higher up in the spine, shoulder, neck, even in the brain. (Flanagan 2000: 15)

Our phenomenology will not tell us when we are wrong about these things. In order to find this out we will need the other disciplines of the natural method. This especially becomes relevant, as we will shortly see, in those cases in which something happens that is counterintuitive.

The splitting auditory attention example was presented to demonstrate that the natural method can help us solving problems concerning conscious experiences. This, then, shows that McGinn is wrong in assuming that there is no method that can solve the mind-body problem; the natural method is a method that he did not consider and it is precisely this method that can solve the problem. In the case discussed in this chapter there were two rivaling hypothesis, and the natural method was used to judge which one was right. Of course, in many cases the method will be used to combine data from the different disciplines, without there being different hypotheses: the natural method just provides us with new knowledge, or is a method to verify of falsify only one hypothesis.

In the chapters seven through eleven I will apply the natural method to a variety of phenomenal experiences. Since this method involves a careful study of the relevant scientific data, the readers interested in the philosophical problem of phenomenal consciousness sometimes have to bear
with me when I present these data, but I can assure them: the philosophical conclusions will follow.

I will start by applying the natural method to experiences of phantom limbs to demonstrate that phenomenal experiences are projected to the (approximate) location of the cause of the sensory input.
CHAPTER SEVEN
CASE STUDY # 1: EXPERIENCES OF PHANTOM LIMBS

In the preceding chapters we saw that philosophical positions like those of Frank Jackson, Thomas Nagel, David Chalmers and Colin McGinn effectively come down to the claim that (contemporary) science cannot solve the mind-body problem. I argued against Nagel, Jackson and Chalmers by refuting their arguments against monism. McGinn – who is a monist – claims that we do not have the cognitive capacities to link concepts referring to the mind to concepts referring to the physical. To undermine this position I presented Flanagan’s natural method, demonstrating that we can find out more about the relation of phenomenal consciousness to the rest of the physical world than McGinn thought possible.

In this and the following chapters I will use the natural method to defend the theory I presented in chapter two. Chapters seven through eleven are five case studies, all focusing on different aspects of the evolutionary shorthand theory of the phenomenal mind.

Chapter eight discusses synesthesia. The upshot there is that phenomenal experiences are presentations that can be used as representations (even as representations of representations). Support for this claim comes from the Charles Bonnet syndrome.

In chapter nine the natural method will be applied to dreaming, demonstrating that we have a natural inclination that phenomenal experiences usually are interpreted as representations. In addition this case study also shows that we are conscious during periods of time common sense has it that we are unconscious. To show that I will also take a look at patients under general anesthesia.

Chapter ten focuses on color experiences. I will demonstrate that colors are secondary properties.

In chapter eleven I will look at the experiences of beauty of faces and look at the evolutionary aspect of the evolutionary shorthand theory. There the upshot is that phenomenal experiences are an economic way of representing.

In the present chapter I will discuss the phenomenon of experiences of phantom limbs, mainly to show that – at least in this case – the phenomenal experiences are projected from within the brain.
to a place outside it. In this way a phenomenal world is generated that is experienced as if it is out there, while it actually is located inside the head.

I will compare my views on projectionism with those of Max Velmans, who defends a projection theory that is fundamentally different. I will start with the phenomenology of experiences of phantom limbs.

1. the natural method applied to experiences of phantom limbs

1.1 the phenomenology of experiences of phantom limbs

Some people have had the misfortune of having one or more of their limbs amputated due to illness, an accident or because they got badly wounded in a war. Even though the actual limb is missing, many of these individuals still have experiences as if it was still there. These illusory limbs are called *phantom limbs*, a term coined by Silas Weir Mitchell. The term is sometimes used to denote a dissociation between the experienced position of a limb and its actual position (Ramachandran & Hirstein 1998: 1604). With respect to the location there are two different kinds of phantom limbs: the phantom limbs that are experienced as being in the same location as the former limb was, and *telescoped* limbs: sometimes patients experience that their phantom hand is directly attached to the stump or even in the stump (Kew et al. 1997: 482) without the rest of the arm in between. However, when tricked in trying to reach with their phantom hand for an object that is a meter away, they do not stand up and walk over to the object, but their phantom hand is zooming in on the object and is no longer directly attached to the stump. This reaching out stops at the normal length of the arm (Ramachandran & Hirstein 1998: 42-43).

What are the experiences these patients have with respect to their phantom limbs? It seems that every kind of experience that one can have in a normal limb can also be experienced in a phantom limb, even though there are important differences between the experiences of individual patients. Most patients report that they can move their phantom limb. These movements are both voluntary, like trying to pick up an object, and involuntary, like trying to ward off blows with their phantom arm (Ramachandran & Blakeslee 1998: 22). There are, however, also cases of patients that do experience their phantom limb in a rigid position and are not able to move this arm. I will return to this in the section on the neurology of experiences of phantom limbs.

Other experiences reported by patients with phantom limbs are experiences of *being touched* (Ramachandran 1993: 10415; Ramachandran & Blakeslee 1998: 29), *feelings of warmth or cold*
The most troubling experience for the patients, however, are not these experiences of having a limb while they actually do not, but are the experiences of pain in their phantom limb. “One specially enigmatic complaint frequently heard from patients is that every now and then the phantom hand becomes curled into a tight, white-knuckled fist, fingers digging into the palm with all the fury of a prizefighter ready to deliver a knockout blow.” (1998: 52) These pains can become so excruciating that some of these patients even contemplate suicide (1998: 22). This has lead to extreme measures: in an attempt to relief the patient of the pain, in many cases the stump was made shorter, based on the belief that the nerve endings at the end of the stump – called neuromas – were irritated, thereby causing the pain. In most cases this did not work. In the discussion of the neurology it will become clear why this radical method indeed cannot relief patients from their phantom pains.

These phenomenological data lead to many questions: how can the pain of phantom limbs be explained?, can the pain be taken away?, and the big question, why are there phantom limbs at all? To answer these questions, we need to apply Flanagan’s natural method, which in this case consists mainly in combining the phenomenology with the neurology of phantom limbs.

In discussing the neurology of phantom limb experiences, two further phenomenological features of these experiences are important. First, the experiences occur almost immediately after the amputation (1998: 23, 34). Estimates vary that between 33 and 75 % of the phantom limb experiences start within the first 24 hours after amputation (Ramachandran & Hirstein 1998: 1605). Secondly, there are reports of patients who were born without limbs that do experience phantom limbs (Ramachandran & Blakeslee 1998: 40-41; Ramachandran & Hirstein 1998: 1604).

1.2 the neurology of experiences of phantom limbs

Ramachandran has developed a neurological theory of phantom limb experiences, that does not only answer the above questions, but is also a theory that shows the neurology behind the projection of bodily phenomenal experiences. The theory has different explanations for the different phenomenal experiences. Central in these explanations is the representation of the body on the cortex: the body image, as discovered by Wilder Penfield.

Penfield (1950) stimulated patients’ brains with an electrode and asked them what they felt. When he stimulated a specific part of the brain, they told him they experienced sensations of the
different parts of their body. In this way Penfield discovered that there is an upside down representation of the body in the brain. This is not an exact 1:1 representation; the lips and thumb, for instance, are disproportionately large. Also, the locations of the different representations of the different body parts are not always where you expect them: the face area is not beneath the neck area, but beneath the hand area. This body image lies in the parietal lobe adjacent to the frontal lobe, hence adjacent to the motor area. This body image is also known as the Penfield homunculus.

The explanation of moving and paralyzed phantom limbs
This body image is both created and maintained by combining information of different sources, for instance, the motor command centers, the muscles and the eyes. If you want to move a limb, the chain of events starts in the motor cortex. The motor cortex sends a signal down to the muscles, and they move. The command that is sent to the muscles is also sent to the body image. In this way the parietal lobe (where the body image is located) "knows" that the limb is supposed to move. The muscles of the limb also send signals to the body image that the movement is actually performed. If the movement is also seen by the organism, then the eyes also send the information that the movement has been performed back to the parietal lobe: the body image in the parietal lobe, then, monitors both the commands that are sent to the body and the response from the moving body that it indeed is executing the commands (or not).

This provides us with an explanation of both the case in which patients experience moving phantom limbs, and the case in which the phantom limb is paralyzed. In the former case signals from the motor command are sent to the parietal lobe. “The commands continue to be monitored by the parietal lobe and are felt as movements. But they are phantom movements carried out by a phantom arm.” (1998: 45)

In the latter case the explanation of the paralyzed phantom arm starts with establishing that the people who have this experience mostly had a long time prior to the amputation in which for example the arm was lying in a sling. The hypothesis is that each time the motor areas did send a command to move to both the arm and the parietal lobe, neither the arm, nor the eyes did sent back signals that the arm was moving. If the patient had this condition for a long time, then the brain might have learned that the arm was paralyzed and the body image might be revised over time. Ramachandran has no exact physiological explanation for this learned paralysis, but
“[w]hatever the physiological explanation turns out to be, when the arm is later amputated, the person is stuck with that revised body image: a paralyzed phantom.” (1998: 46)

**The Explanation of Phantom Touch, Warmth, Cold and Itches**

The explanation of having a phantom limb that itches, feels warm or cold, or seems to be touched also starts with the body image. Tim Pons *et al.* (1991) did an experiment in which four monkeys (*Macaca fascicularis*) had undergone *dorsal rhizotomy*. This is a procedure in which the nerves are cut that send signals from an arm to the spinal column. After eleven years he opened up the skulls of these monkeys and recorded the activity in the body image when different parts of the body were stimulated. As expected touching the hand of the paralyzed arm did not evoke a reaction in the area that normally represents the hand. However, touching the face of the monkey did result in activation of the hand area in the brain (1991: 1859). This suggested that the face area – which lies adjacent to the hand area – had invaded the hand area of the body image. The results were the same for all four monkeys. The question was whether these results could also be replicated in human phantom limb patients. Ramachandran used a cotton swab to stimulate different points on the skin surface of one of his patients who had a phantom hand. He systematically touched the skin of his patients. What he found were two maps – on the actual body – of his patients: a complete map of the patient’s hand on the ipsilateral side of both his face and his upper arm. Pons *et al.* had only found the facial map and wondered why there wasn’t also a map on the trunk (1991: 1859).

A strange phenomenal experience can result from stimulating these two maps at the same time. For patients that have the middle finger amputated, stimulation to the second and fourth finger results also in experiences in the phantom finger. When Ramachandran stimulated the second finger with cold water and the fourth with warm water in a patient who missed her middle finger, “she volunteered that an unusual sensation was experienced – a kind of paradoxical “heat-cold” – as though the phantom was simultaneously warm and cold.” (Ramachandran 1993: 10416)

---

7 Ramachandran used this information to trick the brain of patients with paralyzed phantom arm in thinking the arm is not amputated. He made a box with a mirror in it. The patient puts both his arm and his stump in the box that is so designed that if the patient looks in the box he or she has the illusion of having two arms. If the patient then makes synchronous movement with both arms, in half of the cases the patient immediately experiences movement in both arms.
Based on these findings, Ramachandran formulated the remapping hypothesis of referred sensations. The occurrence of "referred sensations" in the phantom limb [...] is a direct consequence of the remapping [...], which in turn is constrained by proximity of maps in the brain. The reason that there are two clusters of points, for example – one on the face and one near the upper arm – is that the hand in the Penfield homunculus is flanked by one side by the face and on the other side by the upper arm, shoulder and axilla [...]. If the sensory input from the face and from around the stump were to "invade" the cortical territory of the hand, one would expect precisely this sort of clustering of points. (Ramachandran 1993: 10417)

Is there any evidence to support this hypothesis? Yes, there is. The first evidence Ramachandran obtained was from the MEG scans from four amputees. The MEG showed that the maps of the body image indeed had changed as expected, and that the hand and shoulder area had invaded the face area (Ramachandran & Hirstein 1998: 1609).

Further evidence comes from Flor et al. (1995) and Kew et al. (1997). Flor et al. report that of thirteen upper-limb amputees they examined all experienced their phantom limb by touching the face, which was corroborated by MEG recordings showing that activation of the face area also activated the hand area (1995: 482; Ramachandran & Hirstein 1998: 1616). Kew et al. provided evidence that there is a precise 1:1 mapping between the map of the body and the felt location of the sensation in the phantom limb (1997: 2756-2758; Ramachandran & Hirstein 1998: 1616).

So, the remapping hypothesis is well corroborated. Now the following question arises: How do the face and upper arm area invoke the hand area? Ramachandrans suggest that either there is actual remapping and new connections are made between the invading and invaded area, or already existing connections between these areas are no longer disinhibited but activated (Ramachandran & Blakeslee 1998: 34). The fact that phantoms often occur immediately after surgery – estimates vary between 33 and 75% within 24 hours after the amputation – suggest that it have to be old routes becoming active and not new connections, because they cannot be formed by that time. Ramachandran claims that both mechanisms are at work (Ramachandran & Blakeslee 1998: 35). Evidence is now available that at least the first mechanism is at work. Jain et al. (2000) studied five macaque monkeys who had undergone amputation of one of their limbs and found that indeed new connections are made between the face and the hand area.

The explanation of phantom pain

Ramachandran gives four conceivable explanations of why phantom limbs can be painful. Traditionally the nerve irritation theory was developed. While it is not a good theory for the
existence of phantom limbs (there are also subjects who are born with phantom limbs) it might be an explanation of the experiences of pain. The other three explanations all start from the assumption that phantom limbs are (at least partly) experienced due to cortical remapping, and that something went wrong. First, new connections might not only be made to the former hand area in the body image of the parietal lobe, but new connections might also be made to pain centers, so that a touch to the face results in severe pain. Secondly, something might have gone wrong with the volume control mechanisms – the pathways that amplify or dampen input – in the process of remapping. Thirdly, the activity of the touch synapses might not be connected very well, so that they generate junk, which might be interpreted by higher brain areas as pain. Ramachandran does not have a way of choosing between these explanations: “In truth, we really don’t know how the brain translates patterns of nerve activity into conscious experience, be it pain, pleasure or color.” (1998: 51)

2. what experiences of phantom limbs tell us about phenomenal experiences

The first thing we learn from the existence of experiences of phantom limbs is that phenomenal experiences can be present in the absence of the normal physical cause. In the normal situation one experiences a pain in one’s thumb if one actually has a thumb and if there is tissue damage in the thumb. Phantom experiences demonstrate that the same experience can be present while there is no tissue damage, for there is no tissue at all. Still the experienced location is still out there: in the phantom thumb. This of course is also the case in hallucinations, virtual reality, transcranial magnetic stimulation of the brain, direct stimulation of the brain and – as we will see in chapters eight and ten – in synesthesia and in dreams. This shows that when then body image in the brain is appropriately stimulated, a phenomenal world emerges which seems to be out there (outside the head).

The experience seems to be out there, but is it also the actual location of the experience? Some might argue that this question is naïve or that one makes an error of some sort when asking where phenomenal experiences are located. To those I answer in the following manner. Granted that phenomenal experiences are real, we should say what it means that something is real. In the broad physicalist framework it means that something is realized by lower-level (broadly) physical entities, or is a fundamental property. I take it that phenomenal consciousness is a biological property, hence it is not a fundamental property (Cf. chapter one). This means that it has to be (broadly) physically realized. The essence of being physical of physically realized is that one has a location in space, hence this also applies to phenomenal experiences: if they are real, they have to
have a location. *Vice versa* if one says that something is located somewhere, it has to be real, and therefore either belonging to the lowest level of reality or being constituted by it.

Since in phantom experiences we are absolutely sure that there is no physical realization-base out there, phantom experiences show that the experiences actually are located inside the head. Since it seems that the pain is located in the phantom limb, the conclusion is that the phenomenal experience is *projected* to the location out there. Extrapolation of this claim to other experiences leads to the following view. Appropriate stimulation of those parts of the brain that normally generate phenomenal experiences that are used to represent properties of objects, has the effect that the subject has these experience. It does not matter whether the stimulus is the normal property or something else, the result is the same: a phenomenal world that seems to be out there.

There are other experiments that show that experiences are projected to a location outside the head. These experiments involve the projecting of touch to a location that is not occupied by any part of the body of the experiencing subject.

The first experiment is not a real experiment, for we all know the experience and no laboratory is needed. If you are holding a pencil and touch the surface of a table, where do you experience the hardness of the table? As Velmans says “[t]he table feels hard at the point where it is pressed. But there are no sensory organs located at the pencil tip!” (2000: 117) The experience of pressure immediately is projected to the location where the pencil tip touches the table.

The second experiment yields a similar result. Matthew Botvinick and Jonathan Cohen (1998) asked a subject to put one arm on a table behind a screen, so the subject could not see it. Then they placed a visual rubber arm on the other side of the screen in the same position as the arm of the subject. After ten minutes of stroking both the real arm (out of view) and the rubber arm (in view) with a brush, subjects experienced the location of the touch on the rubber arm, not on their own hidden arm (1998: 756). Again, the experience of touch is projected to a location where the subject clearly has no sense organs.

Having said this, my position might seem to be no different from that of Max Velmans (1990, 2000) who also claims that processing of information by the brain leads to a phenomenal world, that is out there. He also agrees that phantom pains are evidence for what he calls “perceptual projection,” the psychological effect of unconscious perceptual processing (Velmans 2000: 116).
However, our views differ fundamentally: I believe that the phenomenal world merely seems to be out there, while Velmans argues it really is out there. This indeed, is the main point he makes. To show where exactly we differ and why I believe Velmans has got this wrong, I will now take a look at his theory, which he calls reflexive monism.

3. reflexive monism

Velmans starts his investigation into phenomenal consciousness not by accepting a theory in advance, but by taking phenomenology seriously. He polarizes the debate about consciousness: either one is a dualist, believing that consciousness is fundamentally different from the physical world, or one is a reductive materialist, which risks eliminating consciousness (Velmans 2000: 108, 172). The first theory, runs against materialist science, the second against our common sense views that we are conscious beings. Velmans wants to take both science and phenomenology seriously. This means that both dualism and reductive materialism have to be wrong. In my opinion this leads to accepting that phenomenal consciousness is a broadly physical property, not reducible to lower-level properties, hence to nonreductive physicalism (Cf. chapter one) As we will see, Velmans thinks that nonreductive physicalism contains a contradiction.

Having polarized the debate, Velmans argues that both the dualist and the reductive materialist share an assumption that is false. The shared assumption is that there is a difference between the perceived objects and our percepts of those objects (Velmans 1990:79; 2000: 104 ff.).

At first sight one might be inclined to agree that the shared assumption is correct, for in our everyday linguistic practices we use the phrase “the perceived object” to denote what we might call the thing-itself, which evidently is not identical to our percepts of the object. The shared assumption, then, amounts to the claim that the perceived object is out there in the world, while the percept of the object is in the head (be it the brain or the mind). Furthermore, it follows that the perceived objects (the things-themselves) are observer independent, while the perceptions of them are not observer independent.

It is against the shared assumption and its consequent claims that Velmans argues. Based on the phenomenal fact that we experience a cat as out there in the world, or a pain as located in the finger, and that we do not have an additional percept of the cat or the pain in our heads, he starts his argument by equating the notions “the object as perceived” and “the percept of the object”; both denote the same. These two notions both differ from “the thing-itself” or “the object as-
described by physical science” (Velmans 1990: 87). Evidently, if one accepts the equation of the perceived objects and our percepts of those objects, then an entirely new interpretation of the shared assumption is the result, an interpretation no one will adhere to, for the shared assumption becomes absurd: something would no longer be identical with itself.

If the perceived object is identical to our percept of the object, the dispute becomes where it is located? Assumed that phenomenal experiences do have a location, they either are where they seem to be, or they are somewhere else. Velmans distinguishes three experienced locations: out-there (e.g. a cat), in the body (e.g. a pain) and in the head (e.g. a thought). (Velmans 1990: 81-82). Let us focus on the first two categories, since there is no dispute about the third category – everyone agrees that thoughts and headaches are in the head (though one might argue about specific examples, like sounds when one listens to music wearing a headphone). Furthermore, if we interpret “out-there” as “not in the head,” then we can ask where the experiences are that seem to be out there: are they where they seem to be, or are they somewhere else?

The answer of the dualist is that they are in the mind, while the reductive materialist claims them to be in the head (Velmans 1990: 93). Even though I am neither a dualist nor a reductive materialist, I agree with both answers, since I believe that the mind is located in the head. Velmans, on the other hand, says that if we take phenomenology seriously, we have to conclude otherwise: the phenomenal world is (out) there where it is experienced (Velmans 1990: 82, 93; 2000: 115; 126; 133).

Though the phenomenal world is (mostly) out there, the neural causes and correlates of this world are in the head. According to Velmans this leads to the following picture. The thing itself is the initiating entity, on which for example light falls, which is reflected. The light stimulates the retina and consequently the occipital cortex is activated. In processing this information a neural representation is created (Velmans sometimes calls this a mental representation, which is not to be confused with a phenomenally conscious state.) In some unknown way, this is projected on the initiating entity whereby the phenomenal object is created out there. The same story goes for bodily experiences. The body is stimulated – say by stabbing the hand with a pin – and the information is processed in the body image. The psychological effect of this information processing is the phenomenal experience of pain in the hand. The neurology of phantom limb experiences shows that appropriate stimulation of the body image results precisely in this projection of the pain. So, whether or not there is an actual hand that is damaged, stimulation of the hand area in the brain leads to the projection of the pain.
Velmans’ main idea is that if we agree that there is a phenomenal world that is experienced as being out there or there where the body is, there is no additional phenomenal world in the head – there is no internal duplicate of the external phenomenal world.

When one gazes at a cat, out there in the world, there is no duplicate experience of a cat “in the mind” [...] or for that matter, “in the brain” [...] Rather, all that one experiences is one cat out there in the world [...] A similar argument applies to one’s own physical body and to all other physical objects out there in the world. (1990: 82)

The reflexive model suggests that in terms of phenomenology there is no actual separation between the perceived body and experiences of the body or between the perceived external world and experiences of that world. (2000: 111)

To the claim that there is just one phenomenal world I can agree, but I believe that phenomenology is wrong about the location: the experienced location of experiences is not the actual location of experiences. This is what phantom pain experiences show: there is no physical object where the pain seems to be located, hence the pain cannot be located there, even though the pain is projected there, hence seems to be there. This is not to change the meaning of the word “pain” from its phenomenological interpretation (the experience) to the physical interpretation (the tissue damage); I am still talking about the phenomenal experience of pain. Claiming that phenomenology shows that experiences of cats and tables are out there is taking phenomenology too seriously. If something has a location, it has to have a physical subvenient base (or it has to be a fundamental property). The liquidity of water is there where the H2O molecules are located, my hand is where the molecules that constitute it are located, etcetera. If the phenomenal world has a location out there, then there also has to be a subvenient or constitutive base for it. It is clear that in the case of phantom limbs such a base is missing. Even though there appears to be a pain out there, there really is no pain out there. So, my objection to Velmans’ reflexive monism is that it confuses appearance of reality with reality. This objection is not new and Velmans has responded to it (2000: 127-131). What is his reaction to this?

Velmans starts his counter argumentation by quoting John Searle (1992) who also made this objection:

Common sense tells us that our pains are located in the physical space within our bodies, that for example, a pain in the foot is literally in the physical space of the foot. But we know that is false. The brain forms a body image, and pains, like all bodily sensations, are parts of the body image. The pain in the foot is literally in the physical space in the brain. (Searle 1992: 63, quoted in Velmans 2000: 12)
I agree with Searle, that common sense has got it wrong. Velmans argues that this is not the case. Scientific data concerning the body image precisely reveal what he claims: direct stimulation of the different parts of the body image do not generate experiences in the head, but tactile experiences located in the corresponding regions of the body (Velmans 2000: 129). This, of course is not an argument against the appearance-reality objection, for the fact that direct stimulation of the brain results in experiences that seem located outside the head, is also coherent with the claim that this is mere appearance. As far as this is meant to support the argument that we have no evidence for the location of sensations in the brain (2000: 129), hence should give up that view, it fails.

Velmans also argues against the appearance-reality objection in the following manner. According to Colin McGinn the fact that brains cause the conscious experiences justifies the claim that they are in the head – if they are anywhere at all (McGinn 1995: 152; Velmans 2000: 130). McGinn’s claim is partly motivated by the fact that one can experience a hand, while the hand is amputated (McGinn 1995: 152). Velmans rightly notices on several occasions in his book that the neural causes and correlates of consciousness are not consciousness themselves, and that we should not confuse these ontological relations. However, this does not establish an argument against the view that conscious experiences are in the head too. In my opinion the fact that both the causes and especially the correlates of consciousness are located in the head makes it plausible that the effects are there too. If nonreductive physicalism is correct, then the neural correlates of consciousness are part of the subvenient base of consciousness, just as the $\text{H}_2\text{O}$ molecules in the glass on my table are part of the subvenient base of the liquidity of the water. If we ask where the liquidity is, then the answer is: in the glass. Of course this is an intuitive argument, and in the case of liquidity there is no projection involved, hence the analogy doesn’t apply in every detail. However, for something to be ontologically real, there has to be a subvenient base, and it is hard to see what that might be other than the brain, in the case of phantom limb experiences. This, then, supports the appearance-reality objection to the claim that the phenomenal world is out there.

Another way to argue in favor distinction between appearance and reality concerning conscious experiences is to show that Velmans is wrong in his claim that there is a contradiction in Searle’s nonreductive physicalism with respect to this distinction. As we have seen, Searle argues that the appearance of the location of pain in the hand is mere appearance: the actual location of the experience is in the head. Hence, phenomenology has got it wrong. This seems to be
contradicted by Searle himself, when he says that “consciousness consists in the appearances themselves. Where appearance is concerned we cannot make the appearance-reality distinction because the appearance is the reality.” (Searle 1992: 122, quoted in Velmans 2000: 131). This seems to be another way of saying that phenomenology cannot be wrong. Velmans argues:

[O]ne cannot both argue that for conscious appearances ‘the appearance is the reality’ and argue that pain which appears to be in the foot is really in the brain, as Searle does […]. In my view, Searle is forced into this self-contradiction by his ‘nonreductive physicalism’. If one takes conscious appearances seriously, one has to accept that pains are not by and large in the brain. But on his version of physicalism, all conscious states are just higher-order (physical) features of the brain, in which case they must be in the brain. As far as I can judge, one cannot consistently hold both positions. (Velmans 2000: 137)

I believe that Velmans is wrong here: one can consistently defend the claims that (a) phenomenology cannot be wrong qua phenomenology, but that (b) phenomenology can be wrong about many other things. Phenomenal experiences are in a sense immune to error: e.g. experiences of pain are always experiences of pain. Furthermore, if the pain is experienced as being localized in the hand, then this also is immune to error, even if there is no hand. More general, experienced locations of experiences are immune to error. However, from the fact that the experience of the location is immune to error, it does not follow that the experienced location of the experience is the actual location of the experience. This is how phenomenology can be wrong.

Let me make the same point with another kind of experience. Accept for the sake of argument that colors are secondary qualities (I will argue for this in chapter ten). If a normal color seeing person under normal circumstances looks at a fire truck she experiences red. No one will deny that she experiences the truck as being red: the red seems to be out there, on the truck. On the secondary quality view of colors the truck is not red at all, hence even though her experience is immune to error, the phenomenology is wrong as far as it represents the fire truck as being red. In chapter eleven I will argue that even though the phenomenology has got it wrong in cases like this, this kind of phenomenal representations is very useful, for it provides short-cuts to the primary properties of objects, in this case those of the fire truck.

In this chapter I have argued that phantom limb experiences show that phenomenal experiences that are located in the head are projected outside the head. Velmans is right that the experienced objects seem to be out there, and that we should equate notions like “the physical object as perceived” with “our percepts of those objects” and not with “objects-themselves.” I disagree, however, that from our experiences of the perceived objects as being out there in the external
world, it follows that our experiences are not in the head. They are in the head: if experiences are ontologically real, they have to have a subvenient base, hence there has to be a subvenient base for phantom limb experiences. In absence of the limbs it is hard to see what can be the external part of the subvenient base, for there is none. This leads to the claim that it is more plausible to assume that phenomenal experiences really are in the brain, and merely appear to be out there. This can be combined with immunity to error, for the experienced external location is immune to error, while the actual location is in the head. Hence, this is no problem for nonreductive physicalism.

In the following chapters I will apply the natural method to other phenomenal experiences, to show why – even though they are in the head – phenomenal experiences seem to be out there.
In this chapter the phenomenon of synesthesia is central. Simon Baron-Cohen and John Harrison (1997a) define synesthesia as follows:

We, along with others […], define synaesthesia as occurring when stimulation of one sensory modality automatically triggers a perception in a second modality, in the absence of any direct stimulation to this second modality. (Harrison & Baron-Cohen, 1997a: 3)

I will start by explaining in more detail what synesthesia is and will use psychological experiments to show that the phenomenal reports of synesthetes are truthful: synesthesia is a real phenomenon that has nothing to do with a mere association of a color with a sound.

The phenomenology of synesthesia suggests that there are two kinds of synesthesia: sensory and conceptual. Psychology supports this distinction and suggests that there is also a distinction between higher and lower synesthesia, where higher and lower refer to the stages in the processing of input by the brain. The neurology of synesthesia shows that the latter distinction is one that has to be made in the conceptual kind. In this case study the natural method is used to establish that we indeed should make the distinction suggested by the phenomenology. The conclusions that follow for the problem of phenomenal consciousness are mainly based on the conceptual kind of synesthesia.

These conclusions are the following. Synesthesia shows that phenomenal experiences can occur in the absence of the regular stimulus. In other words: phenomenal experiences are presentations. Furthermore, these phenomenal presentations can be interpreted and used as representations and even as higher-order representations (without becoming conceptual representations). It will become clear that in synesthesia the usage of the presentations as representations is reliable. From this it will be argued that if synesthetic presentations are used as representations, the phenomenal representations are not misrepresentations.

I will also argue that – at least in the case of synesthesia – the phenomenal experiences and their contents should be located in the head. As we have seen in chapter two (section 2.3) internalism and the secondary quality view of colors go hand in hand. At least in the case of synesthetic experiences, the colors indeed are secondary qualities. From the fact that internalism and the secondary quality view are correct in the case of synesthetic experiences, the suggestion follows
that these views are correct with respect to all our phenomenal experiences. I will argue in chapter nine for internalism and in chapter ten for the secondary quality view of colors.

The reaction to the case study of synesthesia might well be that I should have looked at other and less exotic phenomena as they can build the same case, like dreaming, normal color experiences, direct brain stimulation or drug-induced experiences. First, even though I do agree that there are other phenomenal experiences that establish some of the conclusions I draw based on the phenomenal and scientific data of synesthesia, these conclusions are much clearer in the case of synesthesia. Some of the other examples of phenomenal experiences will be discussed in later chapters though. Secondly, I think that even though the set of phenomenal experiences is heterogeneous, many experiences have more in common than just a what-it-is-likeness. This implies that data about different phenomenal experiences will often lead to the same conclusions. Objections to any choice to investigate one type of phenomenal experiences rather than another can always be made. Lastly, the case of synesthesia is not as exotic as one might expect, rather it is a fairly common phenomenon that is unknown to most people. I will discuss this in the next section after a brief introduction of what synesthesia is.

1. synesthesia

Synesthesia (Greek, syn = together, aisthesis = sensation, perception) is a baffling condition to nonsynesthetes, where stimulus of a sensory modality evokes more than the usual experience. Synesthetes can experiences colors when hearing sounds, or when they smell paint. Some have tactile experiences on their hands when they taste their food, or experiences certain tastes when looking at a scene. Mostly synesthesia is uni-directional and bimodal: e.g., sounds invoke colors, but colors do not trigger sounds, and only colors are the extra qualia. However, sometimes unimodal stimulation leads to more than bimodal experience, like in the case of Luria’s famous mnemonist Shereshveski, for whom sounds triggered experiences in all sensory modules (Luria 1986). It should be noted that synesthesia is not a form of association of – say – a color with a word: if certain words are heard, synesthetes really have experiences of colors.

Synesthesia is a rather unknown phenomenon. It is hardly investigated scientifically, despite the fact that there are many synesthetes among us. Estimates of the ratio vary between 1:25,000 (Cytowic 1995: 33), 1:2,000 (Harrison 2001: 55) to 1:200 (Ramachandran & Hubbard, 2000: 6).
The trait is more common among women than men. Estimates of the women to man ratio vary between 2.5:1 (Cytowic 1989b: 850) to 20:1 (Harrison 2001: 178).

There are no reliable estimates of the adult to children ratio. Cytowic reports of three cases in which subjects claim that they do no longer have the synesthetic experiences which they had as a child (Cytowic, 1995: 35). The disappearance of synesthesia at adulthood supports Lawrence Marks' (1975) claim that "the frequency of synesthesia in children appears to be two to three times that in adults." (1975: 326) The problem with this estimation is that Marks bases it on publications of 1881 to 1923, a time during which there were no good diagnostic criteria. This means that mere association of colors with words might also be categorized as synesthesia. If children are more prone to associate colors with words without noticing that this is strange, then synesthesia would seem to occur more in children than in adults while this is not the case: association is wrongly taken for synesthesia. The doubt about the correctness of the child to adult ratio based on the reports from late nineteenth and early twentieth century is supported by the unlikely high estimates of the ratio of synesthetes to nonsynesthetes in this period, between 6:1 and 8:1 (Révész 1923: 315). Furthermore, at that time synesthesia was also a rather popular game amongst artists, who deliberately made connections between different modalities, while they did not have synesthetic experiences (e.g. Lemley 1999: 84). Classifying these "invented synesthetes" as genuine cases of synesthesia also increased the estimates of the number of synesthetes. Synesthesia is common, but not as common as it was thought around the end of the nineteenth century.

There are two reasons that, even though synesthesia is rather common, it is unknown to most people. Both reasons have to do with how synesthetes evaluate synesthesia. There are two groups. The first group holds the belief that everyone has experiences similar to theirs (Ramachandran & Hubbard 2001: 4), and do not talk about it for this reason. The second group – which seems to be larger given the reports in literature – finds out at a very early age that they have special experiences, which others think odd. “As a child I once mentioned my colors to a teacher, who promptly told my parents I was schizophrenic. That ended my telling anyone about it for quite a while.” (RP in Cytowic 1989a: 49) Hence, they keep quiet about it, because they are afraid that others might think they are mad (Cytowic 1993: 119; Paulesu et al. 1995: 662).

Another reason that today synesthesia is not widely known is the lack of interest of the scientific community until quite recently. Synesthesia was a popular topic in artistic circles at the end of the
nineteenth and the beginning of the twentieth century. At the same time there was a peak in scientific publications on synesthesia. With the arrival of behaviorism the scientific attention faded, which was at its lowest in the 1960s, after which interest increased again. And it is only since the 1990s that we have diagnostic criteria of synesthesia, which allow for clear demarcation between genuine synesthesia, drug-induced synesthesia (Cytowic 1993: 128-129), cross-modal experiences during temporal lobe epileptic seizures (1993: 135), synesthesia that is acquired due to the disease retinitis pigmentosa (Armel & Ramachandran 1999), a heightened association between – say – numbers and colors, and the invented synesthesia of some artists (Lemley 1999). I will discuss these diagnostic criteria in the next section, after discussing the phenomenology of synesthesia.

2. the natural method applied to synesthetic experiences

2.1 the phenomenology of synesthesia

At first sight one might well expect that synesthesia is a condition in which stimulation of one of our five senses would also evoke a sensation in one of the other remaining four, in other words one might expect it to be solely an intermodular phenomenon. However, synesthesia is more than that. First, there are intramodular synesthetic experiences for the visual system: sometimes visually perceived shapes trigger color-experiences.

Secondly, other modules play important roles in synesthesia, for instance, modules that process information about numbers, letters, words, or music. I will start the discussion of synesthesia by looking at its phenomenology: what are the different extra sensations that synesthetes have reported so far, and how can we classify them? I will argue that synesthesia not only comes in different varieties, like sounds evoking colors, or tastes shapes, but that there are (at least) two kinds: sensory and conceptual synesthesia. Recently Grossenbacher and Lovelace (2001) also argued that this distinction should be made.

Sensory and conceptual synesthesia

Let us take a look at our five different senses. At first sight it is conceivable that stimulation of any of the senses can lead to an experience in another sense that is not stimulated at that moment. Many of these couplings are indeed reported. For instance, sometimes smells invoke colors (colored-olfaction): “Muriel still remembers when her father painted the family house. The paint was white. But it smelled unmistakably blue” (Motluk, 1994: 33). Other examples are tactile-gustation (Cytowic, 1993) and colored-gustation synesthesia (Cytowic, 1989a). The following
table shows which of the extra sensations have been reported, and which not. Here, I am concerned only with experiences that are triggered by sensory experiences, whether inter- or intramodular. The varieties of synesthesia that are induced by letters, words and digits are left out for the moment. I should note three things: First, the fact that some conceivable varieties of synesthetic experience are not reported does not of course imply that these do not exist. Secondly, the choice of publications that mention these varieties of synesthesia is arbitrary: the list is longer for each variety that is reported. Thirdly, Sean Day — the president of the American Synesthesia Association — has made a similar list that contains some varieties of synesthesia I did not encounter in the literature. His list is based on interviews with synesthetes that are as yet unpublished.8 (Day, S.A. 2002, personal correspondence).

In addition to our five senses proprioception can be regarded as the sixth sense. There is only one report of synesthesia that might involve proprioception:

> Perhaps the strangest synesthesia is “audiomotor,” in which an adolescent positioned his body in different postures according to the sounds of different words. Both English and nonsense sounds had certain physical movements, the boy claimed, which he could demonstrate by striking various poses. By way of convincing himself of this sound-to-movement association, the physician who described it planned to re-test the boy later on without warning. When the doctor read the same word list aloud ten years later, the boy assumed, without hesitation, the identical postures of a decade earlier. (Cytowic 1997: 21)

All these varieties of synesthesia are of the sensory kind: the experiences are triggered by sensory stimuli. As I just said there is also a kind of synesthesia that seems to be conceptual, for the stimuli that trigger the synesthetic experiences are either spoken or written semantic units. In both cases synesthetes have reported that numbers, letters, parts of words or words evoke colors. This is called grapheme-color synesthesia, and the experiences of color are called photisms. I believe that there are two kinds of grapheme-color synesthesia: sensory and conceptual. In the former case it is a shape or sound that happens to be a word that evokes the photisms, in the latter case it is just the shape or the sound of words that evoke photisms and no other shapes or sounds.

If the grapheme-color synesthesia is sensory, then it is likely that synesthetes might have synesthetic experiences when they hear a number, but not when they see it, or the other way round. Additionally, if a synesthete has these experiences when she sees letters and numbers, the letter O and the number 0 should both evoke the same experience, for it is the shape that...

---

8 This list can be found at Day’s website: http://www.users.muohio.edu/daysa/synesthesia.html.
matters. Also, roman numerals should have the colors (if they have them at all) of the letters not of the Arabic equivalents.

<table>
<thead>
<tr>
<th>visual synesthetic experience</th>
<th>auditory synesthetic experience</th>
<th>tactile synesthetic experience</th>
<th>olfactory synesthetic experience</th>
<th>gustatory synesthetic experiences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>visual stimulation</strong></td>
<td>(A) Cytowic 1989a; Mills et al. 1999.</td>
<td>not reported (*)</td>
<td>Lemley 1999</td>
<td>SD</td>
</tr>
<tr>
<td><strong>tactile stimulation</strong></td>
<td>Lemley 1999.</td>
<td>SD</td>
<td>(C)</td>
<td>Lemley 1999.</td>
</tr>
<tr>
<td><strong>olfactory stimulation</strong></td>
<td>Motluk 1994; Mills et al. 1999; Lemley 1999.</td>
<td>SD</td>
<td>SD</td>
<td>(D)</td>
</tr>
<tr>
<td><strong>gustatory stimulation</strong></td>
<td>Cytowic 1989a; Mills et al. 1999; Lemley 1999.</td>
<td>SD</td>
<td>Cytowic 1989; Cytowic 1993; Lemley 1999.</td>
<td>not reported</td>
</tr>
</tbody>
</table>

Table 8.1 In this table only sensory to sensory synesthesia is listed. (A) The only case in which intrasensory synesthesia is reported is in visual perception. (B) There are reports of people hearing voices, that are triggered by their own voice. (C) Phantom experiences in amputated hands that are triggered by touching the face. These arguably might be called acquired synesthesia, just like the case of the man experiencing colors in retinitis pigmentosa. I am not aware of any reports of (D) and (E) that might be classified as (acquired) synesthesia.

(*) A strange condition is *gaze tinnitus*, in which subjects with damage to their auditory nerve experience ringing sounds when they look to the left (or to the right). This might be called a case of acquired synesthesia, where visual stimuli evoke auditory sensations.

SD: Varieties reported by Sean Day.

If the grapheme-color synesthesia is conceptual, then hearing or seeing a number should both evoke a color. Roman numerals should evoke the same color as their Arabic equivalent. Furthermore, if a conceptual synesthete is “tricked” into thinking about a number, for instance by asking what the sum of two and three is, he or she should also experience the color that is evoked by the number five.
The features of both kinds of synesthesia show up in the studies of synesthesia. Cytowic reports of a synesthete for whom “[r]oman numbers carry the color of the letters that they use” (1989a: 36). Révész describes Judith, a synesthete in which roman numbers evoke different colors than the Arabic numbers (1923: 310). This indicates sensory synesthesia. Révész also reports that the letter O and the number 0 have the same color for this synesthete (1923: 319), supporting my hypothesis.

Some synesthetes report exactly what one should expect if one had conceptual synesthesia. M. – a multimodal synesthete with whom I did some experiments – clearly has synesthetic experiences triggered by concepts of numbers; the Arabic seven has the same color as the roman number and the spoken word “seven.” He also told me that upon being awakened by his alarmclock, he does not experience any extra colors until he has recognized what the time the digital numbers say it is.

If we indeed can make a distinction between sensory and conceptual synesthesia, then this should in the end be based on different neurological causes, not on phenomenology or psychology. I will demonstrate that this indeed can be done.

Making a distinction between sensory and conceptual synesthesia might not be enough to classify all varieties of synesthetic experiences. How should we classify emotions and pains as triggers for synesthetic experiences? Emotions have been reported to evoke colors, for instance that anger is purple (Cytowic 1989a: 34). Also, the feeling of pain is reported both as trigger (1989a: 35) and as response (1989a: 109). A very strange variety is also reported by Cytowic, viz. food is experienced as having numbers. “Food has numbers and falls into “groups,” 3 representing heavy foods, and 1 representing thirsts and lighter foods.” (1989a: 34) These cases of synesthesia do not provide phenomenal data that indicate to which kind of synesthesia they belong, which makes classification hard at this moment. Maybe, future neurological investigation will tell us to which kind of synesthesia each of these varieties belong.

2.2 the psychology of synesthesia

Combining phenomenology with psychology can help us settle the most important question that arises from the reports of synesthetes on their phenomenology: is synesthesia real? There have been a number of alternative explanations of the reports of synesthetes, all claiming that synesthesia is not real. People claiming to have synesthetic experiences sometimes are supposed to talk
metaphorically, like we do when we say that someone wears a loud shirt or that a song is blue. The most important alternative explanation of synesthesia is the learned association theory.

With respect to colored vowels, one possible explanation that comes to mind is that children’s books often present words in colors. Thus, by the association of sound with the printed letter, printed letter with its color, children may tend to associate sound with colors. (Marks 1975: 321)

The learned association theory is a denial of the reality of synesthesia, for it claims that there are no extra experiences, only associations at the level of concepts. However, this learned association theory is very implausible, for synesthetic twins do experience different colors with the same letters and words and publishers of alphabet books make sure that concurrent letters do have different colors, while synesthetes’ reports show that concurrent letters are often different shades of the same color (Baron-Cohen & Harrison 1997). Furthermore, synesthetic children note that colored letters have the wrong color: “My children had a book – Learn Your Numbers – but the colours of the numbers were all wrong.” (SdeM in Cytowic 1989a: 211)

So, the learned association theory is not plausible. But that does not establish the reality of synesthetic experiences. If we take a closer look at the features of synesthetic experiences, we will see that some of them can be used to firmly conclude that synesthesia is real.

Cytowic gives us five diagnostic criteria of synesthesia. Synesthesia is (1) durable, (2) involuntary, (3) projected, (4) memorable, and (5) emotional (1993: 76-79). To have four of these features is enough to diagnose that someone is a synesthete.

In my opinion the fifth criterion that synesthesia has to be emotional is highly problematic. The idea is that synesthetes are very fond of their extra experiences, and that this points towards a link between synesthetic experiences and emotions. However, any experience or cognitive activity can be linked in this manner to emotions. It is widely held that doing logic or math is a rational process. But this does not mean that the experiences that come with them are not emotional. I think that if mathematicians and logicians did not have experiences of beauty and did not greatly care about their work, they would not be doing what they are doing: they are emotionally involved. But if even these prototypical activities of rationality and the accompanying experiences are linked to emotions, it is no wonder that synesthetic experiences are too. So, this criterion is one that picks out every experience, and should not belong to the diagnostic criteria of synesthesia.
Of the remaining four criteria three have been investigated in experiments. The fact that
synesthesia is memorable – i.e. that it can be used for remembering things – has not been
corroborated by experiments. The idea is that synesthetes have an extra way of categorizing those
phenomena that evoke the extra experiences. This has not been corroborated yet, so the reason
for adding this to the list of criteria is based on anecdotal evidence: “She had a green name – I
forget, it was either Ethel or Vivian,” says a woman named Diane. She confuses the actual names
because they are both green, but she remembers the synesthetic greenness.” (Cytowic 1993: 77)

The three remaining diagnostic criteria are confirmed by experiments. I will now turn to these
features, after which I will discuss the psychological data that support the distinction between
conceptual and sensory synesthesia.

**Criterion # 1: synesthetic experiences are durable**

Synesthesia is **durable**, which is to say that the experiences are consistent over time. Because of
this, synesthesia can be investigated scientifically since experiments can be **repeated**. Durability can
be used to establish the reality of synesthesia. In order to do so, we have to find cases where it
has been established that people that claim to have synesthetic experiences indeed score better on
a test than control subjects. This indeed has been done. Baron-Cohen *et al.* (1993) used a list of
130 words that was read to synesthetes with colored speech perception and to a non-synesthetic
control group. Contrary to the nonsynesthetes, the synesthetes were not warned in advance that
they would be retested at a later date. Both groups had to name the color of the words evoked.
(The control group had to associate colors with these words of course.) Baron-Cohen *et al.* report
that “[o]verall, 92.3% of the responses from the synaesthetic group were consistent one year
later, whilst only 37.6% of the control group were consistent, even after one week.” (Baron-
Cohen *et al.* 1993: 422)

**Criterion # 2: synesthetic experiences are involuntary**

The fact that synesthetic experiences are involuntary can also be used to show that synesthesia is
real. This can be shown with a test derived from the Stroop-test. J.R. Stroop (1935) was
interested in two questions:

1. Will there be a difference in performance time between (A) reading words that name colors
   when these words are printed in black ink, and (B) reading the same list of color words
when they are printed in colors different from the ones they name (e.g. the word “red” printed in green ink)?

And

(2) Will there be a difference in performance time between (A) naming the colors of the ink in which nonsemantical items like squares are printed, and (B) naming the colors of the ink in which color words are printed when there is a mismatch between the color of the ink and the color word (e.g. the word “red” printed in green ink)?

The results were that there is no significant difference in the first case: whether the color words were printed in black or in a color different than the ones they named made no difference. In the second test, however, the subject response time increased by 74% when they had to name the color of incongruently colored color words compared to the list of squares in the same color. This is now known as the Stroop-effect.

Mills et al. (1999) and Odgaard et al. (1999) used Stoop-like tests to confirm that grapheme-color synesthesia is real. I used a similar test to show that the photisms – the extra experiences of color – M. experiences when he sees numbers are genuine. This test goes as follows.

The subject: M. is a multimodal synesthete, who experiences colors with written and spoken numbers and words, shapes and music. This indicates that he has both conceptual and sensory synesthesia. He also experiences tastes as colored and shaped. Two nonsynesthetic subjects were used as controls.

Stimuli: Prior to the experiment M. was asked which numbers evoked which colors. These are: 0 white; 1 beige-grey; 2 blue (between Prussian blue and ultramarine); 3 red (deep red but slightly lighter than wine-red); 4 yellow; 5 blue (lighter than 2, more sky-blue); 6 red (a bit like 3, but a little more dull); 7 purple; 8 dark brown-grey; 9 grey; 10 light yellow. He experiences the color of each digit whether it appears alone or in a larger number (with the exception of 10). These information was used to make lists of numbers used as stimuli.

Eight different lists of thirteen items were made for six conditions: (a) colored X’s to determine the baseline; (b) a list containing normal words to determine whether there is influence of
semantic items on the reaction time; (c) a list containing incongruently colored color-association words (e.g. “sky” printed in green or “grass” in red ink); (d) incongruently colored color words (the Stroop-stimuli); (e) a list of numbers printed in colors that are incongruent compared to the experiences M. has; and (f) a list like (e), but where each item on the list is printed in the color of the experience the preceding item evokes, e.g. “555” evoking a experience of blue, is followed by “999” printed in blue. Since in reading a list one tends to look at what is coming next, there is a double stimulus to say “blue” when one has to name the color of the ink in which “555” is printed. All items were printed in 18-points Arial.

**Procedure:** All lists were tests in which one had to name the color of the ink of the items. The lists were presented face down and in random order. On the commando “one, two, three” the lists had to be turned around, after which the colors of the thirteen items had to be named followed by “stop”. Time was measured using a stopwatch.

**Results:** The reaction times of the control subjects were as expected. There was no difference between lists (a), (b), (e) and (f). Compared to this baseline reaction time their times increased with 19.1 % for list (c) and 38.7 % for list (d).

The reaction times of M. also were as expected. His reaction time was the same for lists (a) and (b) which is the baseline. Just as in the case of the control subjects his reaction time went up for lists (c) and (d), respectively 16.1 % and 51.6 %. His reaction time for list (e) was almost the same as his Stroop-reaction: 52.4 %. In the case of list (f) containing incongruently colored numbers each followed by a number in the color of the experience, his reaction time even increased with 78.8 %.

From this increase in reaction time, we can conclude that M.’s photisms are involuntary and genuine. At least for the part of his synesthetic experiences that come with seeing numbers his synesthesia is real.

**Criterion #3: synesthetic experiences are projected**

So far we have seen that the first two diagnostic criteria can and have been be used to establish by means of psychological experiments that synesthesia is a genuine phenomenon. Recently it has been shown by Smilek *et al.* (2001) that synesthesia indeed is – as some synesthetes report – *projected.*
When shown a digit, the color of the photism is not experienced “in the mind’s eye,” but rather it is experienced as though the color was “externally projected” onto the digit – for example, when C., a digit-color synaesthete whom we have studied […] is shown a black 4, she describes the blue photism triggered by the 4 as being seen “out there, on the page, overlaid on top of the 4.” (2001: 930)

Smilek et al. designed an experiment based on the assumption that if the experiences indeed were projected onto the digit, this influences her visual perception of the digits. They briefly showed C. a black digit against either a congruent or an incongruent background. C. indeed performed better in naming the number in those situations where the background was incongruent to her color-experience of the digit. This shows that C.’s experiences are real and that the photisms are projected onto the digits (2001: 931). It also shows that in the case of C. first the normal information is unconsciously processed and then the extra experience is generated and projected onto the target. C.’s conscious awareness of the digit follows this projection. This is contrary to M. who first has to recognize the digit and then has the extra experience (Cf. supra). I will later argue that these are two subkinds of conceptual synesthesia.

**Conceptual synesthesia**

The phenomenology of synesthesia indicated that it is plausible that we should distinguish sensory from conceptual synesthesia. Though the evidence should come from neurology, a psychological experiment indicates too that the distinction should be made. If there is such a thing as conceptual synesthesia, then indirect triggering of the concept also should evoke the color while this should not be the case in sensory synesthesia. Dixon et al. (2000) indeed found this in C. They presented C. with sequences consisting of the following elements: a digit, an operator (like plus or minus), another digit and a color patch (either congruent of incongruent to the experienced color of the solution to the problem). C. then had first to name the color of the patch and then the solution to the arithmetic problem. The hypothesis was that, “[i]f automatic photisms can be induced simply by calculating the solution to arithmetic problems, then they should interfere with C.’s ability to name the colour of the patch on incongruent trials but not on congruent trials.” (2000: 365) This indeed was the result they found, which lead to the conclusion that an external stimulus (for example, a physically present numeral 7) is not required to trigger a photism. Rather, activating the concept of a digit by a mental calculation was sufficient to induce the colour experience. Thus, although C.’s photisms are both consistent and automatic, they do not require a physical inducing stimulus to elicit them. (2000: 365)

**higher and lower synesthesia**
Ramachandran and Hubbard (2001) also investigated grapheme-color synesthesia. They believe that there are two subtypes which they call higher and lower grapheme-synesthesia, related to the higher and lower cortical areas that process sensory input. Visual input is processed in a serial manner (though of course that takes place via parallel routes) starting at the retina, then going to the lateral geniculate nucleus, to V1 and then to the more specialized areas like V4 (see for more details chapter ten). The areas early in the processing of visual input are the lower areas, the areas that come later in the process than higher areas. Depending on what area is involved in synesthetic experiences, the synesthesia is classified as lower or higher synesthesia.

They did some experiments showing that in two of their grapheme-color synesthetes the induced colors are sensory in nature, that is triggered by the shape and not by the concept. They base this conclusion on the fact that the induced colors lead to pop-out. If we are presented with a computer screen filled with (digitally written) two’s and one five, then it is very hard to spot the five. The colors the synesthetes experiences are different for the five and for the two. The five pops out for these two subjects. “We measured their performance and found that they were significantly better at detecting the embedded shape than non-synesthetic control subjects[...], making it clear that they were not confabulating and could not have been ‘faking it.’” (2001: 7)

There is another experiment which supports the thesis that the colors these synesthetes experience are sensory rather than conceptually in nature. If a digit is presented in the periphery of the visual field it can still be recognized. However, if it is flanked by other digits, this becomes more difficult. So, the number five can be recognized in the periphery of our visual field, but if it is at the same place in our visual field and if it is surrounded by other numbers we no longer are able to recognize it. In this task the synesthetes could not immediately identify the target digit, but were able to infer from the color they experienced which digit was flanked by other digits (2001: 8).

These experimental results clearly show that the experiences of these synesthetes are triggered by the sensory input, not by the concepts correlated with them. These synesthetes did not experience the colors of the numbers, when confronted with roman numerals (2001: 8), suggesting the same.

Ramachandran and Hubbard also encountered grapheme-color synesthetes that did experience the same colors for the Arabic and roman numerals. This leads them to postulate a distinction
between higher and lower grapheme-synesthesia. At first sight the distinction between higher and lower synesthesia seems to be the same as the distinction between sensory and conceptual synesthesia. As it will become clear in the next section the former is a distinction between two subkinds of conceptual synesthesia.

2.3 the neurology of synesthesia

Before demonstrating that the suggestions from phenomenology and psychology are indeed correct and that we should distinguish between sensory and conceptual synesthesia and between lower and higher synesthesia, I first will discuss the controversy between the different neurological theories that are supposed to explain synesthesia in general.

There is a dispute between (1) the theory that states that synesthesia is a normal limbic feature that surfaces only when the cortex does not inhibit it, and (2) the theory that says that synesthesia is caused by cortical cross-wiring – a breakdown of the modularity of the brain. Cytowic defends the former theory, while Baron-Cohen and Harrison, Ramachandran and Hubbard, and Smilek et al. defend the latter theory. Though these theories initially seem incompatible, I believe that both may be right.

Neural theory # 1: synesthesia takes place in the limbic brain

Richard Cytowic (1993) tells the story of his discovery of synesthesia in Michael Watson. Cytowic was invited to dinner, and when Watson tasted some of the food he was preparing he said that there weren’t enough points on the chicken. Watson tasted shapes. Furthermore, Watson always drank some coffee in the morning, which resulted in his synesthetic experiences being less vivid than normal. Drinking beer made his experiences more vivid. Since caffeine is a cortical stimulator and alcohol a cortical inhibitor, Cytowic concluded provisionally that Watson’s synesthesia had its origin in the limbic system.

Cytowic places synesthesia in the limbic system of the left hemisphere; to be precisely in the hippocampus. For this, he gives two reasons. First, “there are persons with hippocampal epileptogenic foci who have synesthetic experiences relatable to a seizure, but who are not synesthetic otherwise.” (1989a: 174)

The second reason, an anatomical one, is that the hippocampus is a place in the brain where signals come together “from functionally different and geographically independent areas.” (1989a: 174)
I think these reasons indeed support Cytowic’s hypothesis. I take it that hippocampal epileptogenic foci causing synesthetic experiences clearly point to the hippocampus as playing an important role. The second reason, however, is quite controversial. It is intuitively very appealing that we should look for a place in the brain where different kinds of information come together, so information can be combined into one informational state or experience. As we will see in chapter nine, it is by no means clear that there is such a place (even though the hippocampus is a place where signals come together it is not the place where binding occurs.)

The evidence for Cytowic’s theory comes from brain imaging. A PET-scan of Watson’s cerebral blood flow (CBF) was made and it showed something very unusual. The baseline CBF was so low that his cortex looked almost dead. While experiencing synesthesia, this got even worse. Instead of an expected increase in blood flow, the average blood flow in his left hemisphere was eighteen percent less than his baseline. When given amyl nitrate (the drug poppers) Watson’s synesthetic experiences were enhanced: instead of feeling just one glass column while tasting spearmint, he now felt there were hundreds of them. What amyl nitrate does, is to cause a shut-down of the cortex for about ten seconds, thereby enhancing the relative activity of the limbic system. This, then, is Cytowic’s direct evidence for his claim that synesthesia is a normal feature of the limbic brain, which is normally inhibited by the cortex (Cytowic 1989a; 1993).

Neural theory # 2: synesthesia is caused by cortical cross-wiring

Not everyone is convinced that Cytowic’s theory is correct. Baron-Cohen says in an interview with synesthete Alison Motluk that “[w]e would be cautious about generalizing about that [the single case study of Michael Watson].” (Motluk, 1994: 36) I do not think this is a reason to dismiss Cytowic’s direct evidence: even if just one man has such strange CBF, it is highly relevant, for it shows that synesthesia is physically possible without cortical activity. Harrison and Baron-Cohen, on the other hand, seem not convinced and present an alternative theory. This theory also comes with some direct evidence.

In short, Harrison and Baron-Cohen claim that “where as in nonsynaesthetes audition and vision are functionally discrete, in individuals with synaesthesia a breakdown in modularity has occurred.” (1997b: 119) This means we have to look for neural connections between known visual and auditory areas (in the case of colored-hearing). There are two theories of how this crosswiring might work: the local crosswiring theory and the disinhibited-feedback or reentrant theory.
The local crosswiring theory

The local crosswiring theory says that there are extra connections between modules that lie adjacent to each other, which normally are not directly connected. In the case of grapheme-color synesthesia the theory goes as follows. The fusiform gyrus in the left hemisphere contains the lower color-processing areas V4, V8 and the visual grapheme area. The latter lies adjacent to V4. Nobre et al. (1994) found out that the posterior fusiform gyrus is active when subjects process words and groups of letters forming nonwords. The anterior portion of the fusiform gyrus is only activated by words that do have semantic content. Neither parts are active when processing other stimuli like faces or butterflies.

The angular gyrus of the left hemisphere is necessary for abstract numerical calculation. Damage of this area leads to acalculia: the patient cannot do arithmetic anymore. Adjacent to the angular gyrus lies the superior temporal gyrus, which contains the higher cortical color-processing areas.

Ramachandran and Hubbard believe that local crosswiring can explain synesthetic experiences. Their hypothesis is that there are extra neural pathways between (1) the visual grapheme area and the lower color-processing areas in lower grapheme-color synesthetes and (2) the angular gyrus and the superior temporal gyrus in higher grapheme-color synesthetes.

If there indeed are such extra pathways between normally separated areas, this might be explained by failure of pruning (Ramachandran & Hubbard 2001: 11). There is evidence (Maurer 1993: 111) that the cortex of young infants is not so specialized as in the adult. For instance sounds evoke responses in the temporal cortex (which is normal), and in the occipital cortex (which in adults is the visual area). This suggests that neonates are synesthetes. The idea is that certain pathways are pruned during development. In synesthetes this pruning might fail, leaving pathways between two brain areas intact, leading to synesthetic experiences.

The disinhibited-feedback or reentrant theory

Grossenbacher and Lovelace (2001) do not think that synesthetes have local crosswirings between – say the angular gyrus and the higher color-processing areas – but instead they believe that the neuroanatomy of synesthetes is the same as that of nonsynesthetes. The difference can be explained by looking at how their brains work. In normal brains some multi-modal sensory areas receive signals from different pathways and then send them on to the area the signal came
from (feedback) or to another area. If inhibition of this convergence area does not work properly, signals (say from a shape-processing area) that were not meant to be send forward to another area (say a color-processing area) activate that area, resulting in synesthetic experiences.

Grossenbacher and Lovelace do not say what specific brain areas they have in mind, for they propose a general theory of the neurology of synesthesia. Smilek et al. present a similar theory in which they do address the specific neuroanatomy of grapheme-color synesthesia (in what I believe to be a conceptual synesthete). In conceptual grapheme-color synesthesia, the mechanism might be the following. First a number is presented in black ink.

Information regarding form activates shape-processing areas in the primary visual cortex [...] as well as the extrastriate areas in the lingual and fusiform gyri (human V4). Form, comprised of the line segments in a digit, is then processed by anterior fusiform and PIT areas where the form activates the meaning of that digit[.] [...] We believe that for synesthetes like C, PIT activation of the meaning of a digit influences color processing in V4 via feedback connections. (2001: 933)

This disinhibited-feedback theory or reentrant model differs from the crosswiring of Ramachandran and Hubbard, in that crosswiring occurs via normal pathways that are not inhibited, while Ramachandran and Hubbard postulate extra neural pathways between brain areas that normally are not connected.

There is evidence that might show that either one of these cross-wiring theories is right. Again, the direct evidence comes from brain imaging. The only problem is that the direct evidence is not evidence of the existence of any pathways, but of the fact that synesthesia is a cortical process. The direct evidence presented here, is indirect evidence for the occurrence of these pathways, hence we cannot decide which of the two cross-wiring theories is correct (Cf. Harrison 2001: 15-19).

Paulesu et al. (1995) made PET scans of six synesthetes who experienced spoken words as colored.

In both groups [six synesthetes and six nonsynesthetic controls] word stimulation compared with tone stimulation activated the classical language areas of the perisylvian regions. In synaesthetes, a number of additional visual associative areas, including the posterior inferior temporal cortex and the parieto-occipital junctions, were activated. [...] Synaesthetes also showed activation in the right prefrontal cortex, insula and superior temporal gyrus. [...] These results suggest that colour-word synesthesia may result from the activity of brain areas concerned with language and visual feature integration. (1995: 661)
Here the evidence is undeniably in favor of the theory that places the cause of synesthesia in the cortex as opposed to Cytowic’s theory that places it in the limbic system. Given the results the synesthetes involved are likely to be all conceptual grapheme-color synesthetes. This is supported by the report of Paulesu et al. that these synesthetes did not experience colors with nonsemantic sounds (1995: 663).

**Sensory and conceptual synesthesia**

There are many varieties of synesthesia if we classify the synesthetic experiences by triggering module and extra quale. These varieties are often treated as varieties of the same kind. I believe that this is mistaken, and that this mistake lies at the root of the dispute between the cortex theory and the limbic system theory. I believe that the evidence on both accounts is strong, which leads me to conclude just one thing: there are (at least) two kinds of synesthesia, one that has its cause in cortical cross-wiring, and one that has its cause in the limbic system in combination with cortical inactivity. Furthermore, I believe that this classification corresponds to the distinction between sensory and conceptual synesthesia.

If we look at the two brain imaging studies, Cytowic clearly investigated a sensory kind of synesthesia: tasting shapes. It seems undeniable that his limbic theory is correct here. In any case, it cannot be a cortical phenomenon.

Paulesu et al. investigated a cases of grapheme color synesthesia, which suggest that they investigated conceptual synesthesia. I believe that we should (at least) distinguish three varieties here. First of all there is the variety in which forms are the trigger of colors. These forms are only forms that are associated with semantic content, hence it is not merely a variety of form to color synesthesia, which would be a variety of sensory synesthesia. The explanation could be that there is crosswiring between (a) the posterior fusiform gyrus, which is activated by only these kind of forms, whether they do have semantic content or not and (b) the lower color processing areas. This then is a variety of lower grapheme-color synesthesia. A similar account should be given for the kind of grapheme-color synesthesia that has spoken language as trigger, which seems to be the variety Paulesu et al. investigated: “[o]ur subjects did not have synaesthesias during reading, unless they subvocalize what they read.” (1995: 666)

---

9 There even might be a third kind of synesthesia. Some people report seeing bright colors when they heard sudden loud sounds. This might be related to specific nuclei in the hypothalamus and brainstem. “Neurons belonging to these nuclei appear to fire whenever something important or salient occurs, such as a loud noise, a flash of light, or a sudden pain.” (Edelman and Tononi 2000: 46) It might be that if there is some crosswiring in these systems with some parts of the visual cortex, one might experience colors when a loud noise is heard.
For both the second and third variety of conceptual synesthesia the trigger is the semantic content, either of words or of numbers. The explanations might be that there is either a crosswiring between (a) the anterior fusiform gyrus, which processes only those words that do have semantic content, and (b) the lower color processing areas (the second variety), or that there is crosswiring between (a) the angular gyrus, which processes numbers, and (b) the higher color processing areas (the third variety). These would then respectively be varieties of lower and higher grapheme-color synesthesia.

<table>
<thead>
<tr>
<th>sensory synesthesia</th>
<th>conceptual synesthesia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Explanation:</strong> Normal limbic feature, with failure of cortical inhibition.</td>
<td>The form of the grapheme is the trigger.</td>
</tr>
<tr>
<td>lower grapheme-color synesthesia</td>
<td>The concept of the grapheme is the trigger.</td>
</tr>
<tr>
<td><strong>Explanation:</strong> crosswiring between posterior fusiform and lower color processing areas</td>
<td>higher grapheme-color synesthesia</td>
</tr>
<tr>
<td><strong>Explanation:</strong> crosswiring between anterior fusiform gyrus and lower color processing areas</td>
<td>crosswiring between angular gyrus and higher color processing areas</td>
</tr>
</tbody>
</table>

**Table 8.2** Suggestion for classifying synesthesia in kinds and varieties, based on phenomenology, psychology and neurology.

3. **the relevance of synesthesia to the problem of phenomenal consciousness**

What is the relevance of synesthesia to the problem of phenomenal consciousness? First of all, synesthesia shows that the content of (at least) these phenomenal experiences is located in the head. I will present the argument for this in a moment. This argument will lead to other important conclusions: since it is evident that concepts are not colored the synesthetic experience that is evoked by a concept does not directly represent any primary property of the concept. This might be seen as a case of misrepresentation. I will argue that synesthetic experiences are not misrepresentations, but that they are merely presentations. Furthermore, since these presentations are time and again evoked by the same concepts, the presentation can reliably be used by the synesthete to represent the concept. If this is done, the presentation is used as a representation of a representation, for a concept itself is already a representation. This, in turn, shows that a first-order theory of phenomenal consciousness cannot be right, for it cannot account for all phenomenal experiences.
3.1 externalism about intentional content

Synesthesia shows that externalism about intentional content cannot be correct. Adam Wager (1999) also argues for this conclusion. He takes representationalism (whether it is concerned with wide or narrow content) as the claim that phenomenal character is identical to representational content. This is the strong version. The weaker version of representationalism expresses a supervenience relation: phenomenal character supervenes on representational content. Wager rightly argues that Tye adheres to the strong interpretation (Cf. chapter two, section 2.7). We also have seen that Tye combines this with an externalist view of content. I also defend the strong version of representationalism (for those phenomenal presentations that are used as representations) but combined with internalism.

Wager argues that synesthesia poses a problem for Tye’s externalist theory for his

theory locates the phenomenal content of experience in the world, rather than in states internal to the individual. It is because of this that Tye’s theory cannot account for cases of extra qualia, cases in which subjects with different internal states have experiences with different phenomenal contents though their environment is held fixed. (Wager 1999: 268)

Let us take the following example. A nonsynesthete has an auditory phenomenal experience when he hears middle C. A synesthete has a different phenomenal experience consisting of the auditory experience in combination with an experience of – say – red. Based on the phenomenology of synesthesia Wager takes it that these two experiences should be treated together (1999: 270) because for color-hearing synesthetes experiences of colors are part of what-it-is-like to hear. Hence, there is a difference between the synesthetic and nonsynesthetic phenomenal representation of middle C. If both a synesthete and a nonsynesthete represent the same external feature – middle C – and the phenomenal content lies in the external world, then externalists like Tye have a problem. According to externalism the intentional content is middle C in both cases, which would imply (according to strong representationalism) that the phenomenal character should be the same. This is not the case, if we take the synesthetic experience together with the auditory experience.

I do not believe that this is the way to show that externalism is false, for it is based on what I believe to be a false supposition viz. that the phenomenal representations of middle C are different for a synesthete than for a nonsynesthete. Even though I agree with Wager that externalism about phenomenal content is incorrect, I believe that separating the synesthetic from
the nonsynesthetic experience seems to provide a way out for the externalist: the auditory phenomenal experience of middle C of both the synesthetes and nonsynesthetes have the same content. The externalist now only has to give an additional representational account of the extra red-experience the synesthete has when hearing middle C. However, I believe that in the end this is no solution for the externalist, for the account of what is represented by the extra experience has to be an internalist account, which then conflicts with the externalist account of the auditory experience. Let me explain.

If we treat the two experiences of a synesthete separately, and we claim that the synesthetic experience is not a representation of the direct cause of the regular experience, we need to answer the following questions.

1. Is the extra experience a representation? If so, what is it then, that is represented by the extra experience?
2. If the extra experiences is not a representation, why does the phenomenology tell us that it is (e.g. a color is evoked by a number, hence the number seems to be represented as being blue)?
3. If answers to these questions can be given, why do these answers imply that externalism about intentional content is incorrect?

I will answer that (1) synesthetic experiences are presentations that can be used as representations, (2) that transparency explains the phenomenology, and (3) that externalism about intentional content is incorrect in the case of synesthesia for the extra experiences evidently show that the secondary property view of colors is correct (which implies internalism, *Cf. supra*).

Those who defend both the claim that all phenomenal experiences are representational and that externalism is correct seem to have a way out: synesthetic experiences are misrepresentations. In order to defend my answers to the first three questions then, I have to answer a fourth question:

4. Why are synesthetic experiences not misrepresentations?

*Question 1: is the extra experience a representation?*
Let us take a look at conceptual synesthesia and assume that some synesthete experiences blue when hearing, seeing or thinking about the number four. The options we seem to have are the following:

(a) The synesthetic experience is not a representation at all;
(b) The synesthetic experience is a representation representing (a property of) the number four;
(c) The synesthetic experience is a misrepresentation that seems to be representing (a property of) the number four.

Suppose (b) is the correct view. What then is the synesthetic experience of blue a representation of in the case of the conceptual synesthete experiencing blue when she sees, hears or thinks about the number four. We have seen that for conceptual synesthesia the cross-wiring theory is the best explanations: in conceptual synesthesia there is an extra neural route from the brain area that is necessary for the conceptual representation to a color area in the brain. For example, the conceptual representation of the number four in the angular gyrus also activates the color area in the higher color processing areas resulting in the experience of blue. This suggests that it is the entire concept “four” that in turn is represented by the experience of blue. The same conclusion follows from the phenomenology of conceptual synesthesia.

What is evident (both to synesthetes and the rest of the world) in the case of conceptual synesthesia is that the number four is not blue in the sense that a car can be blue. In the case of a blue car, common sense has it that being blue is one of the many properties of the car. If the synesthetetic experience is representational, then it does not represent a property of the number four, but rather represents the number four itself. This is supported by the fact that for conceptual synesthetes two plus two might equal yellow (Dixon et al. 2000) where yellow then can be used as a synonym for four.

If we take synesthetic experiences as representations, then we should say the following in the case of conceptual synesthesia. In a synesthete a concept (e.g. “four”) triggers an extra experience (e.g. blue). This extra experience then is a representation of that concept. The concept itself is a representation, hence the synesthetetic representation is a representation of a representation.
The first consequence of this is that – at least in this case – colors are secondary properties, since concepts evidently are not colored. Since internalism and the secondary view of colors go hand in hand, this also supports internalism, suggesting that internalism might also be correct in other cases.

Secondly, the synesthetic experience is a higher-order representation since it is a representation of a representation. Synesthetic experiences are phenomenal experiences, hence they are also nonconceptual representations. So, synesthetic experiences are both higher-order and nonconceptual representations at the same time. This has a major implication for theories of phenomenal mind: first-order representational theories that take higher order representations as conceptual – like those of Tye or Dretske – cannot be right for it will leave out synesthetic experiences.

Normally (phenomenal) representations are not represented again. What happens in the case of synesthesia, then, can also be put as follows. A representation of a number triggers an unusual activation in a synesthete, causing her to experience a color. This experience is merely a presentation, just as a drug-induced experience would be, or one induced by an epileptic seizure or migraine. It is evident that in none of these cases there is a property that the phenomenal experience is a representation of. A synesthete knows that the number four is not colored (at least not in the way cars are colored).

This leads just to one conclusion: the phenomenal experiences in these cases are not representational in nature. By this I mean that they are not necessarily or intrinsically representational. That does not mean that these presentations cannot be used as representations. (Just like a stone is not a tool in nature, but can be used as a hammer.) Synesthetic experiences can be used as representations. Ramachandran and Hubbard (2001) did the following experiment (Cf. supra). If one is presented with a number on a piece of paper in the periphery of one’s visual field, one can see the number. However, if it is surrounded by other numbers this becomes very hard. Synesthetes can use their extra phenomenal experience in this particular experiment to infer which number is surrounded by other numbers. If a synesthete experiences the number four as blue she knows that it is a four that is in the periphery of her visual field when she experiences blue. Here, the extra phenomenal presentation is used as a representation: blue represents four. A synesthetic experience of a color then is a presentation that is used to represent the whole
concept, and not just a property of it, as is the case in which a car is represented as blue. In that case blue represents a property of the car, not the whole car.

Evidently, synesthetic presentations that are used as representations are *indirect* representations: four is not represented as four by the synesthetic experiences, but as blue. This then, shows that even though (a) is in principle the right view, (b) is correct as soon as synesthetic experiences are used as representations. (I will refute (c) in a moment.)

I believe that this theory applies to most phenomenal experiences. Phenomenal experiences are merely presentations: they are not representational *in nature*. However, they can be used as representations. Synesthesia, then, provides a very clear example of this claim. Before I turn to the objections I will first answer questions two and three.

**Question 2: why are synesthetic representations projected through the represented representation?**

Now, what about the second question? If phenomenal experiences are representations of representations, why is it that middle C seems to be red, and not some brain state, concept or other phenomenal experience? Put metaphorically, I would say that since phenomenal representations are transparent, the projected redness goes right through the auditory property and gets projected on the middle C out there.

We do not experience the location of our phenomenal experiences inside the head (with the exception of headaches and the like). If we see an apple, its redness seems to be out there. This also goes for the other experiences evoked by the apple: the experience of something round, hard and sweet out there. If one of these phenomenal representations in turn would trigger another phenomenal experience, and this experience is interpreted as a representation, then evidently one will not interpret it as the representation of a representation, for one is not aware of the representation it is an representation of: hence one will interpret the extra experience as a representation of the object that evoked the normal experiences.

**Question 3: how does this amount to internalism?**

How does synesthesia show that externalism is false? At first sight synesthesia seems to be in support of externalism. First, we already saw that the externalist might claim that the intentional content of the auditory experience in both the synesthete and the nonsynesthete is the same (e.g.
middle C.) The phenomenal auditory experience in both cases is also the same, so there is no problem here: both the representations and their contents are the same.

Secondly, the intentional content of the extra phenomenal experience of the synesthete is the auditory phenomenal representation of middle C, which – in principle – is an externalist account of the content of the representation. (Even though the representation of middle C is in the head, it is not part of the synesthetic representation itself: the content of this extra experience is external to the representation.) This then, here is an externalist account of the two representational states and their content.

The problem for the externalist account of representational content is that it goes hand in hand with a primary view of (e.g.) colors (Cf. chapter two, section 2.3). In normal cases such a view seems to be plausible: it seems that it really is the ripe tomato that is red and that sugar really is sweet. But I already argued that this is not true that the auditory representation of middle C is red. The same goes for every other phenomenal representation that is phenomenally represented by a synesthetic experience. The most evident case is that of the higher subkind of conceptual synesthesia. Concepts do not have any properties like being colored, tasteful or smelly. How then, can an externalist give an account of this within a primary view of color, tastes and smells? How can one say that the concept “four” has the primary quality of being blue? I think the only way to do this is to regard synesthetic experiences as misrepresentations. This brings us to the objections to my answers. They all can be summarized as follows: if synesthetic experiences are misrepresentations, none of the above conclusions follow. Before addressing this problem, let us look in more detail at the possible objections to my answers.

### 3.2 possible objections

First I will present three objections. These objections stem from different positions but lead to the same conclusion: if it can be shown that synesthetic experiences are misrepresentations, then none of the above conclusions follow. Secondly, I will give two criteria that have to be met for something to be a misrepresentation. Thirdly, I will answer all three objections at once, by showing that synesthetic experiences do not meet these criteria. This leads to the conclusion that these extra phenomenal experiences can be viewed as presentations or as representations but never as misrepresentations.
Objection # 1

I have argued that synesthetic experiences are presentations that can be used as representations: they are not representational in nature. This then goes against radical representationalist theories that claim that every phenomenal experience is a representation.

Radical representationalists have to argue that the synesthetic experiences indeed fall in the same category as the color experiences evoked by drugs and artificial stimulation of the brain. They should claim that in all these examples there is something that seems to have a color. Since there either is no object at all (e.g. the hallucination of a pink elephant) or the object does not have the experienced property (e.g. the number four printed in black ink evoking an experience of blue), these representations are misrepresentations. In this way a representational theory of phenomenal consciousness could stick to the claim that phenomenal experiences are representational, without having to accept that there are experiences that are not representational in nature but merely presentational.

Evidently, if synesthetic experiences are misrepresentations the secondary property view does no longer follow. Nothing about the ontological status follows from misrepresenting a four as blue: there simply is no property that is represented by the experience, hence no argument in favor of the secondary property view of colors can follow (nor for the primary quality view).

Objection # 2

Those in favor of FOR theories might object in a similar way to my claim that synesthetic experiences can be used as higher-order nonconceptual representations. They might claim experience of blue is not a representation of a representation, but of the thing out there: the number four printed in black ink. Again, since the number four is printed in black ink, representing it as being blue is a misrepresentation.

Objection # 3

If synesthetic experiences would misrepresent the thing out there, then the conclusion that externalism is incorrect does not follow. First, we already saw that the secondary property view does not follow if synesthetic experiences are misrepresentations. If the secondary property view does not follow, evidently it cannot function in an argument against externalism.
Secondly (the argument is similar to that showing that the secondary property view does not follow), if the experience is a misrepresentation, then there is no property or object out there to which it reliably corresponds, hence nothing follows about properties out there (or properties in here).

So radical representationalism, first-order representationalist theories of the phenomenal mind, and externalism about phenomenal content can be defended against my conclusions based on the data of synesthesia if synesthetic experiences are misrepresentations. It is time I address this problem.

**Two criteria for misrepresentation**

For me, representation is the same as reliable tracking. If I represent a square object as being square, I reliably track this property of that object. This is direct representation, for it represents the primary property of being square as it is. The same goes for the representation of whole objects. If I represent a cat as a cat, then I reliably track this particular animal. In both cases it is evident that successful tracking only can occur if there actually is an object a property of an object. Furthermore, these representations are reliable because they represent the property or the object as it is, resulting in appropriate actions of the representing organism.

The same goes for indirect representation: this also is reliable tracking, implying that there is something that is tracked and that is done reliably so, resulting in the appropriate action of the representing organism. Suppose for the moment that colors are secondary properties (I will demonstrate this in chapter ten). On the secondary property account of colors, ripe tomatoes are represented as red but are not really red. There is a primary property (or a set of primary properties) that is responsible for the experience of red in us. This property is indirectly represented, and it is also a reliable representation. That it is a reliable representation of course does not lie in representing the property as it is, as is the case in direct representation. It is a reliable representation, for the combination of the following reasons. First of all, under most normal circumstances tomatoes with this property always evoke experience of redness in us. Secondly, we do not get into trouble when action upon it: eating red tomatoes is the appropriate reaction to experiencing red when seeing a tomato in normal circumstances.

For reliable representation, the circumstances have to be normal. If they are not normal, one might make mistakes, for instance when one is almost blind or if there is almost no light, a cat
may be mistaken for a rabbit, or a rotten tomato may seem to be red instead of brown. These are cases of misrepresentation. Of course, in normal situations one might also make mistakes, for instance when one thinks one sees someone familiar in the street in broad daylight, when this actually is a stranger.

If representation is reliable tracking, then misrepresentation is unreliable tracking. In other words: if something is misrepresented, it is tracked, but not reliably so. This brings us to the two criteria of misrepresentation. First, there has to be something either an object or a property of an object that is misrepresented. The second criterion is that the tracking is unreliable: a rat is represented as a cat. The fact that it the tracking is unreliable, surfaces when one acts upon the experiences, for instance when one tries to stroke the rat – because one thinks it is a cat – and it bites, or when one takes a bite of the rotten tomato. These two criteria give us a tool to show that synesthetic experiences are not misrepresentations.

In short, to misrepresent something $A$ as $b$, there has to be (a feature of) something that is misrepresented in a particular case. The same goes for indirect representations, which are not misrepresentations: It does not follow from the representation of $A$ as $b$ that $A$ is misrepresented. Something is an indirect misrepresentation only if it does not represent $A$ as $b$, but as $c$, e.g. the grass is not being represented as green but as blue.

**Synesthetic experiences are not misrepresentations**

Synesthetic experiences meet the first criterion. We have seen that a synesthetic presentation is used as a representation, it is used as a representation of a representation. Hence, it is used to represent – say – the concept four as represented in the brain.

Synesthetic experiences fail the second criterion. An important feature of synesthesia is that the extra experiences are consistent through time. Furthermore, synesthetes do not experience any problems that might be the result of synesthesia in dealing with the world. This contrary to cases of misrepresentations, like the case where a rat is misrepresented as a cat. Even if someone would always misrepresent rats as cats, he would find out that he is, for he would have problems in dealing with the (natural) world: he would try to stroke a rat, which will be reprimanded by the rat. Typically, synesthetes do not have any problems dealing with the world as a result of their extra experiences. On the contrary, synesthesia seems to provide an extra way of (correctly) categorizing the world, synesthetes mostly claim to benefit from their experiences as it comes to
dealing with the world. It is precisely because their extra experiences are consistent through time, that they can reliably use their extra experiences as representations. Many synesthetes that experience sounds also as a color can tell tones apart by tone, but also by color. This might help them in tuning a musical instrument: the experience of red tells that this is an E. Synesthetic experiences, then, fail the second criterion, for if they are used as representations, this is done reliably so.

In conclusion, synesthetic experiences – when used as representations – are not misrepresentations. These experiences are presentations that can be used as indirect, nonconceptual representations of representations. These experiences are projected onto the cause of the normal representation, resulting in the interpretation of that cause as having the extra secondary property. I argued that it actually is a secondary property of the normal representation.

This might seem a very bold conclusion, but I believe it is not. If I am right, all phenomenal experiences are indirect, nonconceptual representations. Hence, this should also be the case for synesthetic experiences. And if all phenomenal representations are representations that represented secondary properties, so will synesthetic experiences.

Two suggestions following from synesthesia

The implication for externalism about phenomenal content is that – at least in the case of synesthesia – it is wrong: the normal representation has a secondary property. As shown in chapter two, this is inconsistent with externalism, hence it has to be incorrect. If the conclusions about synesthesia can be shown to follow from normal phenomenal experiences, then externalism about phenomenal content in general would be incorrect. So, the first suggestion that follows from the discussion of synesthesia is that externalism and the primary quality view of colors are incorrect in general. In chapter ten I will show that this indeed is the case.

First I will apply the natural method to dreaming in chapter nine. This case study also leads to the same suggestion that internalism and the primary property view of colors is correct. Furthermore, this case study proves a second suggestion that follows from the discussion of synesthesia: we have a natural inclination to interpret presentations as representations even in those cases we know that the presentations are not representations. Synesthetes know that numbers are not
colored, still they claim that numbers are experienced as being colored. In the next chapter I will argue that we all have this natural inclination.
In this chapter I will apply the natural method to dreaming, mainly to show that we have a natural inclination to interpret phenomenal presentations as representations as suggested by the prior case study. Our mind-brain is a device to interpret the presentations caused by sensory data and it just is not turned off during sleep. This results in what we call dreams: we interpret all the phenomenal noise that is generated during sleep as representations. If it is indeed the case that our interpretation device is not shut off during sleep, then we have to be conscious during the whole period of sleep. This is a counterintuitive claim: most people will say that they do not dream constantly during sleep. Furthermore, it is widely accepted that we only dream during Rapid Eye Movement or REM sleep.

Owen Flanagan himself has applied the natural method to dreaming in his book *Dreaming Souls* (2000). Parts of this chapter draw on this book. I will apply his method to demonstrate that his claim that we are always conscious during our life is more plausible than one might expect at first sight. At least in the case of sleeping he seems to be right. That is enough for my purposes. Flanagan’s claim is stronger: we are conscious during all our life, hence also in cases of general anesthesia. I will demonstrate that there are patients that were conscious during general anesthesia, but that we cannot conclude that every patient is.

The claim that we have the natural inclination to interpret all phenomenal presentations as representations is also supported by the Charles Bonnet Syndrome, a phenomenon that will be briefly looked at in this chapter.

A conclusion following from the natural inclination claim is that phenomenal experiences and their contents are located in the head. Since internalism and the secondary property view of colors go hand in hand, this cases study suggests the same as the case study of synesthesia: colors are secondary properties. This issue will be addressed in chapter ten.

1. **the phenomenology of dreams**

The investigation of dreams starts with its phenomenology. We all dream, and dreams seem to come in two different kinds. To help classify these two kinds of dreams, we can use Flanagan’s
distinction between (1) single mental events and (2) mental episodes (2000: 55). Single mental events are events that do not have a theme or a plot, while mental episodes do. Seeing a coin on the pavement, or an error message on your computer screen are single mental events. Within the class of mental episodes Flanagan distinguishes between (2a) narrative structures, and (2b) narratives. The experience of fixing a flat tire, giving someone directions have a long enough narrative structure to be called mental episodes. However, for mental episodes to be a narrative something more is needed than simply having a narrative structure. What also is needed is that episodes are connected to each other into a larger story.

“Some dreams are narratives.” (2000: 57) These are the first kind of dreams: the bizarre dreams containing vivid imagery in which for instance people might change into other people or even plants, while apparently remaining their identity. Dreams that are narratives are always about people and things you know or about fictional persons and objects, never about real people or things you do not know. (I cannot dream about your friends that are unknown to me.) To this extend dreams are self-expressive. The phenomenal fact that dreams often are narratives is important, for this already shows that during the night our brains go on doing what they do during the day: make sense of our experiences. This will be relevant in the discussion of the evolution of dreams.

The second kind of dream is the worrying dream. Dreams belonging to this kind are not narratives, but are more than single mental events. We can experience dreams that do not seem to end: in your sleep your are constantly worrying about some futile thing, like did I put my shoes under the table or under the chair? You keep repeating this question, and for some reason this problem is very troublesome: the dream usually is a nightmare. These kind of dreams are less vivid than the bizarre dreams and have a more thought-like structure. (Note that phenomenology might differ from person to person, in this case resulting in different experiences of dreams. Some people seem to have no familiarity with the second kind of dream.)

A phenomenal aspect about both kinds of dreams is that we mostly do not remember what our dreams were about or even that we have dreamed at all. This phenomenology gives rise to the following suggestion: if we are so extremely bad at remembering that we did dream, then maybe we’re dreaming all the time while asleep. Flanagan believes that there is good reason to believe that we are always conscious while alive, hence always conscious while asleep, hence constantly dreaming while asleep.
The view I am going to assume is that we are always, while alive, conscious. If it is true that while alive we are always conscious to some degree, and if it is true that we are always alive and not dead when we are asleep, then it is true that we are always conscious while asleep. We are always dreaming while asleep. (2000: 68)

Taken together, the questions that rise from this phenomenology of dreams are the following.

1. Can we explain why dreams seem to come in two different kinds?
2. Can we explain why we are so terribly bad at remembering dreams?
3. Are we always dreaming while asleep? (Are we always conscious while alive?)

Before I will see how Flanagan uses the natural method to answer the questions, I will add another question he poses.

**A philosophical-biological question about dreams**

In philosophy of mind we are interested in the relation of the mind to the body and the environment in which it evolved. How are mental states realized? How do mental and other broadly physical states influence each other? Why did they evolve? In discussing Flanagan's application of his method to dreams I will not only focus on the questions that arose directly from the phenomenology, but also on the following question which Flanagan categorizes as a problem of philosophy of mind and biology (2000: 20).

4. Is dreaming functional?¹⁰

Answering these questions can only be done if we use the extended version of the natural method. In turn I will discuss the neurology and evolutionary biology of dreams. These provide most of the data that we need to answer the above questions. Furthermore, in answering these questions an important aspect of the evolutionary shorthand theory of phenomenal consciousness results: we have a natural inclination to interpret presentations as representations, even if there are no objects that are represented.

**2. the neurology of dreams**

---

¹⁰ Here, 'function' has to be understood in an evolutionary context: dreams have no evolutionary function. This does not imply that dreams are epiphenomena tout court for they clearly are not: we report about them, interpret them, write books about them. Dreams do have causal powers.
The neuroscientific data can provide us – in combination with the phenomenology – with the answers to the first three questions. I will address these in turn. To answer the first question, we will have to take a look at our sleeping cycle and at which parts of the brain are stimulated during the different stages of this cycle.

The usual EEG records frequencies of 30 Hz and an amplitude of 50 microvolts in awake persons. This changes when one falls asleep. A typical night of sleep starts with an episode of nonREM or NREM sleep. This is followed by an episode of REM sleep. In a normal adult this cycle takes between 90 and 100 minutes, after which it starts again.

NREM sleep is divided into four stages. Stage one is characterized mainly by its brain wave pattern, which is in the \emph{theta} frequency band (4-8 Hz). In this stage the eyes make slow rolling movements, mostly just before the \emph{alpha} (8-12 Hz) and \emph{beta} waves (12-18 Hz) change into the theta waves. This stage is mostly very brief (ten to fifteen minutes for normal young adults). The theta rhythm is irregular and slow; the voltage of these waves slowly increases. In this stage one is easily awakened.

Stage two is characterized by the appearance of two distinct wave forms. (The background activity of stage two is similar to the activity in stage one: theta waves). The first of these wave forms is called the sleep spindle. In normal young adults the frequency is between the three and eight spindles per minute. The second wave pattern is the K-complex. These are sharp negative spikes followed by a little slower positive component. The frequency is about three per minute. They can either occur spontaneously or they can be provoked by stimulation of the senses, for instance with a loud bang. Even though one is not as easily awakened such a loud bang can of course wake a person.

Stage three and four form what is called the slow-wave sleep, characterized by \emph{delta} waves, which have a higher amplitude than the waves of the first stages (between 75 and 200 microvolts), but a slower frequency (0.5-4 Hz). Stage three is a deeper stage of sleep that stage two (a louder noise is needed to wake one up). In this stage the spindles and K-complexes may also occur, but the stage is characterized by the high amplitude and the slow frequency. Stage four NREM sleep also is characterized by the slow-wave sleep, without the spindles and K-complexes. This is a deeper sleep than stage three. So, the brain waves pattern of the awake state (high frequency and low
voltage) is reversed (low frequency and high voltage). After stage four we rapidly go back through
the third and the second stages of NREM sleep, after which we enter REM sleep.

REM sleep is the state in which most of our vivid dreams occur. This stage is characterized by (1)
rapid eye-movement & (2) ponto-geniculo-occipital waves (PGO waves). These PGO waves
originate in the pons. Just as in the waking stage the brain waves of the REM sleep are low
voltage fast-activity waves unlike those of NREM sleep, which are high voltage and low-activity.
The REM stages takes between 20 and 30 minutes.

When we are awake the lateral geniculate body of the thalamus (G) receives input form our retina
after which these go further to the visual processing areas. When we are in REM sleep stimuli
originate in the pontine brainstem (P) that are send to the lateral geniculate body and the occipital
cortex (O), which is the main visual processing area of the brain. Flanagan concludes that “This
much accounts for the saliency of visual imagery in the dreams of sighted people. [...] This
strongly suggests that dream imagery operates on the same anatomical areas as does awake
perception.” (2000: 82) So, here we encounter a stage in the sleep cycle that correlates with our
vivid and bizarre dreams, where neurology provides us with an explanation of why the dreams in
this stage are so vivid.

This also supports Flanagan’s claim that we dream during NREM sleep. If the dreams that occur
during REM sleep are vivid in imagery and we can explain this via the activity of the visual areas
in the brain, then it is plausible that the dreams that are not vivid occur during some other
stadium of sleep, which has to be NREM sleep. Of course the claim that we dream during
NREM sleep is highly controversial.

If we take a look at the firing rate of the individual neurons and not at groups, then there is not so
much difference between waking and both REM and NREM sleeping. It is not the firing rate but
the firing pattern that makes the difference between sleeping and waking. While asleep the
neurons in the brain fire in large numbers at the same time. This in contrast to the waking states
where different groups of neurons “dynamically assemble and reassemble into continuously
changing patterns of firing.” (Edelman & Tononi 1999: 72) This implies that during NREM sleep
the number of neural states is drastically reduced. If one believes that we only dream during REM
sleep this might indicate – as Edelman and Tononi think it does – that for conscious states neural
activity is required “that changes continually and is thus spatially and temporally differentiated.”
This is not Flanagan’s conclusion, he argues that dream scientists focus too much on the vivid imagery of the REM dreams, thereby ignoring the thoughtlike NREM dreams.\footnote{“One hypothesis I have for why many researchers underestimate NREM dreaming, even to the point of thinking it is not dreaming, has to do with overrating the importance of robust sensory imagery, especially visual imagery in dreams. Once sensory imagery is taken as definitive of dreaming, then dreaming is equated with mentation during REM sleep, and the sensory dull, but thoughtlike, mentation of NREM sleep is ignored. This then leads to the mistaken conclusion that NREM sleep, especially stages 3 and 4 NREM sleep, is a period of unconsciousness.” (Flanagan 2000: 83)}

Flanagan presents another reason for believing that we dream during NREM sleep, viz. that sleepwalking and sleeptalking occur during NREM sleep (2000: 84). These phenomena indicate that the sleeping subject has conscious experiences.

I will present some evidence in the last section of this chapter that might make Flanagan’s claim that we are always conscious while alive more plausible. For the moment, I will address the issues about dreams and accept that we indeed do dream during both REM and NREM sleep. The neurological explanation of the two kinds of dreams, then, is that they occur during different stages of sleep, in which different brain areas are activated, corresponding to the dream-experiences one has.

The second question that arose from phenomenology can be answered via neurology: why is it that we are so bad at remembering dreams? The answer lies in the chemicals that are either produced or not produced during sleep. Flanagan uses Alan Hobson’s AIM model to show what happens during sleep with the activation level of our brain waves (A), the input source (I) and the modulation level (M), which says what neurotransmitters are active in the brain. In the case of sleeping and waking two kinds of neurons and neurotransmitters are of interest: cholinergic and aminergic neurotransmitters. Cholinergic neurons synthesize the neurotransmitter acetylcholine, which plays a key role in the generation and regulation of REM sleep. Aminergic neurotransmitters synthesize amongst others serotonin and norepinephrine. These both are important for attention, learning, memory and reasoning.

When we are awake (A) there is high activation of the brain, (I) the most important input comes from external stimuli, and (M) aminergic neurotransmitters like serotonin and norepinephrine dominate over acetylcholine (2000: 80; 89).
During NREM sleep (A) brain activation is low, (I) the stimuli of brain regions comes from other regions, and (M) there is a balance between the aminergic and the cholinergic systems (2000: 80).

In REM sleep (A) the activation of the brain again is high, (I) the stimuli are just as in NREM internal and (M) cholinergic takes over from aminergic. Neurons releasing serotonin and norepinephrine turn off during REM sleep, while acetylcholine is on (2000: 89).

So, what happens during sleep is that the release of some of the neurotransmitters that are involved in memory is turned off. “This helps explain why memory for dreams is degraded.” (2000: 89-90)

We have now the answers to the first two questions: (1) there are two different kinds of dreams, for there are (roughly) two different kinds of brain states during sleep corresponding to the different types of dreams, and (2) we are so extremely bad at remembering dreams, for during sleep the levels of neurotransmitters needed to store information are too low.

Though I am skeptical about the truth of Flanagan’s claim that humans are always conscious while alive, it can be shown to be more plausible than we might think at first sight. The answers to the preceding questions support – but do not prove – this claim. If the widely held belief that we dream only during REM sleep was correct, then obviously Flanagan’s claim would be wrong. Furthermore, the fact that we are bad at remembering our dreams shows that we dream more than we can remember, which suggests that Flanagan’s claim might be right.

It is an unproblematic claim that during the moments that we are awake we are conscious. We have experiences. However, we tend to forget most of them. You might recall what you did have for breakfast yesterday and what program you watched on television, but you probably do not know that for the same day a year ago. So, forgetting what we were conscious of is not such a strange thing when large periods of time are between the conscious event and the day one tries to remember things. Furthermore, it helps that the fact one tries to remember is a relatively unimportant one that occurred on an average day. One is more likely to remember things that were special itself or occurred on a not so typical day.

Not only long periods of time make it hard to remember what one was conscious of; just try to remember a complete sentence of the previous page. Even though you were conscious of this a moment ago, recollection is impossible.
This should at least make it plausible that we forget or do not explicitly store many of the things we were conscious of, a claim which is implicit in the claim that contrary to our common sense beliefs we are always conscious during our lives. This should take away the intuition that Flanagan’s claim is evidently false. In the following section I will present neurological data that show that this claim can even be supported. The idea is the following. First, we need to establish a plausible neural marker of consciousness. Secondly, we should find this marker during NREM sleep (in accordance with NREM dreaming). Furthermore, this marker should be found during other periods of which common sense believes that these are periods of unconsciousness, like general anesthesia. In this way neurology might tell us something about our phenomenology.

3. the 40 Hz thesis

Francis Crick and Christof Koch (1990/1997) have suggested that the neurons involved in conscious experiences fire synchronously at a rate of 40 Hz. These oscillations, called gamma rhythms, actually range from 35 to 75 Hz, but are commonly referred to as 40 Hz oscillations. “We shall refer to these neuronal oscillations as the 40 Hz oscillations, though the frequency on any one occasion is not very precise and can vary between 35 and 75 Hz.” (1990/1997: 286) The hypothesis is that

neurons which represent the same object or event might fire their action potentials in temporal synchrony with a precision in the millisecond range. However, no such synchronization should occur between cells which are part of different representations, or cell assemblies. (Engel & Singer 2001:18)

Crick and Koch presented their 40 Hz thesis as a possible solution to the problem of binding visual experiences. If we take a look at just the visual cortex, then we see that different areas process different features:

neurons in the area MT are mostly interested in motion and depth, those in area V4 in color and shape, and those in 7a in position in space relative to the head or the body. So far no single area has been found whose neurons correspond to everything we see. (1990/1997: 282)

Why do we not experience a red blob, a movement, and a certain shape, but do we see a red car in front of us?

It seems likely that at any moment any individual object in the visual field is represented by the firing of a set of neurons. This would not cause any special problem if the members of the set were in close proximity (implying that they probably interact somewhat), received somewhat similar inputs and projected to somewhat similar places. But because any object will have different
characteristics (form, color, motion, orientation) that are processed in different visual areas, it is highly reasonable to assume that seeing any one object involves neurons in many different visual areas. The problem of how these neurons temporally become active as a unit is often described as "the binding problem." As an object seen is often also heard, smelled, or felt, this binding must also occur across different sensory modalities. (Crick & Koch 1990/1997: 284; Crick 1994: 208)

The suggestion of Crick and Koch is that the solution to the binding problem lies in gamma band synchronous oscillations.

Koch and I [...] suggested] that this synchronized firing on, or near, the beat of a gamma oscillation (in the 35- to 75-Hertz range) might be the neural correlate of visual awareness. (1994: 245).

They based their assumption on experiments done in cats where synchronous oscillations were found in the olfactory cortex and in the visual cortex and wide apart as 7 mm. These were found both in cats that were lightly anesthetized and cats that were alert.

At the time that Crick and Koch presented their hypothesis, there was not much experimental evidence that supported their speculative proposal. Is there now evidence that gamma band synchronous oscillations indeed are closely correlated with consciousness? Kevin Sauvé (1999) presents a brief overview of experiments that show that this indeed is the case. I am not concerned here with the question what the exact relation is of these oscillations to consciousness: are they identical to consciousness, are they the cause or are they the supervenient base of consciousness? What we want to know is whether there is a strong correlation, which might indicate any of these relations. If we find this correlation, then we have a good reason to consider GSO to be the neural marker of consciousness.

Sauvé presents the recent evidence from microscopic recordings: recordings that do not record large groups of neurons as is done in EEG or MEG (where at least one million neurons are needed, and one cannot be sure whether there is a synchronous firing of 40 Hz or that the measurement is a result of other collective firings). He reports for instance that a cat performing a task in which it had to look at a cue before some action had to be performed (which was rewarded) the correlated firings “displayed highly correlated firing among pairs of neurons in both hemispheres of the visual cortex, as well as in parietal cortex and frontal motor areas.” (1999: 214)

Also binocular rivalry experiments have been used to see whether the 40 Hz thesis is right. In binocular rivalry a different stimulus is presented to each of the eyes, for instance lines moving up to the left eye and lines going down to the right. The phenomenal experiences are either of
lines going down or of lines going up. These percepts switch in a manner similar to our perception of the front or back of a necker cube. Increases in synchrony and regularity of gamma band oscillations were found in the visual areas of cats when one particular visual percept dominated. During that domination the suppressed stimulus exhibited a decrease in synchrony and regularity in the gamma band oscillations (1999: 214).

Not only are synchronous gamma band oscillations found in visual areas during a visual task, they were also found in the motor areas during bimanual manipulations where synchronous activity occurred between neurons that even were 8 mm apart (1999: 215).

Sauvé also presents additional evidence based on EEG and MEG recordings. GSO were found in (1) the auditory thalamocortical region during experiments where subjects had to distinguish between one or two auditory presented “clicks”, (2) the somatosensory cortex during experiments where subjects had to distinguish between one or two “taps” presented to a single index finger; (3) the visual cortex during experiments where subjects had to distinguish between a left- or right-facing Dalmatian in a black-and-white speckled picture; (4) the left hemisphere in experiments where subjects had to distinguish between real and pseudowords, where activity decreased in the latter case; (5) the motor cortex contralateral to the finger used in a task where subjects intentionally had either to move their finger as fast or as slow as possible (1999: 217-221).

This body of evidence leads Sauvé to conclude the following: “these experiments strongly suggest that GSO are causally related to stimuli discriminations and motor outputs that typically occur as part of our paradigmatic cognitive state: consciousness.” (1999: 222)

Andreas Engel and Wolf Singer (2001) also present a list of recent evidence for the same claim. After evaluation of this evidence they conclude that

By far the largest proportion of investigations have been devoted to the visual system of either cat or monkey [...]. Precise synchronization [...] occurs in a large number of different neural systems and across a wide range of species. It has been observed in all sensory systems, in the motor system and in memory/association structures. Species include primates, carnivores, lagomorphs, rodents, birds, reptiles, amphibia and insects. Also shown is the fact that, in many studies, synchrony between separate neurons is accompanied by gamma-band (> 20 Hz) occurs. (2001: 20)

While this does not prove the 40 Hz hypothesis beyond any doubt, it does corroborate it well enough to justify further investigation. So, let us assume that 40 Hz oscillations indeed indicate
conscious experiences. Can we find these oscillations in deep sleep, which should be the case if both (1) Flanagan’s claim that we are always conscious while alive, and (2) the 40 Hz thesis are correct? Indeed, we can.

Rodolfo Llinás and Urs Ribary report that “40-Hz activity occurs in an organized fashion [...] during the dreaming state.” (1993: 2078) They also report that during deep sleep these oscillations are reduced but not absent, meaning that “the amplitude of these oscillations differs from that of wakefulness and REM sleep.” (1993: 2080) Also Steriade, Amzica and Contreras (1996) have found that GSO are present during slow-wave sleep in cats: “far from exclusively characterizing the brain-states of waking and REM sleep, fast oscillations are also present during the behavioral state of resting sleep or under general anesthesia with ketamine and xylazine.” (1996: 425) So, here we have the first neurological evidence that points in the direction of Flanagan’s psychological hypothesis that the phenomenological reports of thoughtlike, worrying dreams occur during NREM sleep.

The fact that we have found support for the thesis that we might have conscious experiences during NREM sleep, does support the thesis that we are always conscious during sleep. It does also support – but to a lesser extend – the stronger thesis that we are conscious during all our lives. Common sense has it that besides (deep) sleep there are also other states in which we believe that we are not conscious, like general anesthesia. If Flanagan’s stronger claim has any credibility, we should also find evidence that we are conscious when we are under general anesthesia. As follows from the quote of Steriade et al. above, GSO have been found in cats that where anesthetized with ketamine and xylazine. The interesting question is whether we can correlate such findings back to the phenomenology and psychology of general anesthesia in humans.

The normal phenomenology of anesthesia is that we are not aware of anything during general anesthesia. This of course, is what the anesthetic agents are supposed to accomplish. However, it is not unusual for patients to recall aspects of events that occurred during surgery. Moerman et al. (1993) report of twenty-six patients that had conscious experiences during general anesthesia. The most frequent experiences were auditory, followed by experiences of paralysis and pain.

Since, auditory experiences are the most common, the auditory area’s in the brain might be best to look for the GSO during general anesthesia if we want to correlate phenomenological with
neurological data. This is what Schwender et al. (1994) did. They tested both specific and nonspecific anesthetic agents. A tape was played during the operation.

The tape had a duration of approximately 10 min. The text contained mainly positive suggestions for postoperative recovery. In this context a short version of the story of “Robinson Crusoe” was told. It gives an example how a person overcame a very difficult situation in his life. The story was meant to be a parable for the patients of how to deal with their present difficult situations and to facilitate postoperative recovery. (1994: 131-132)

During the operation activity of the auditory cortex was recorded. Three to five days after the surgery they were tested for implicit and explicit memory. The results were that in the case of the nonspecific anesthetic agents during surgery the 40 Hz oscillations were either reduced or completely suppressed. No implicit or explicit memory of the tape was found.

In the case of the specific anesthetic agents, the “40-Hz oscillatory activity is preserved during general anesthesia.” (1994: 136) However, none of these patients had explicit memories of the contents of the audiotape. When control subjects – who did not hear the tape – were asked for their association with “Friday” none of them reported anything about Robinson Crusoe, while of the 30 patients who did get to hear the tape seven replied “Robinson Crusoe.” Looking at the brainwaves in the auditory parts of the brain patients without implicit memory did not show the 40 Hz oscillations while the tape was played during their operation, while patients with implicit memory did show 40 Hz oscillations at that time.

This proves that sometimes patients are conscious during general anesthesia and that this is correlated with GSO in the auditory cortex. However, patients do not explicitly remember that they were conscious. In other cases neither the experiences nor the GSO were present, which suggests that anesthetic agents can do what they are supposed to do: eradicate consciousness. This would imply that Flanagan’s claim that we are always conscious while alive is too strong. It might be the case that the patients without the explicit memory of auditory conscious experiences did have other kinds of experiences and GSO in the relevant areas. However, for the moment it seems more prudent to adhere only to the weaker claim that we can have conscious experiences during periods that are commonly regarded as periods of unconsciousness.

12 The specific anesthetic agents that bind to receptors or defined brain regions were: benzodiazepines, midazolam, diazepam, flunitrazepam, fentanyln and ketamine (1994: 135-136). The nonspecific anesthetic agents were: isoflurane, enfurane, thiopentone, etomidate, and propofol (1994: 133).
So far the natural method has proved itself useful. We now know why there are two different types of dreams, why we are bad at remembering dreams and that we do have many more conscious periods in our lives than we remember. Furthermore, the 40 Hz GSO are a reliable marker of consciousness.

4. the evolutionary biology of dreams

Do dreams have an evolutionary function? Flanagan argues that dreams are *the spandrels of sleep.* “Spandrel [...] is an architectural term that refers to the triangular space left over when arches are placed next to each other at right angles [...]. Spandrels are inevitable by-products of arch and column design.” (2000: 104) The evolutionary analog holds for dreams Flanagan claims: they are evolutionary by-products of sleep, which was selected for. Dreaming is just as much a by-product as the sounds hearts make or the color of our bones. Having a heart that functions properly also means that one has a heart that throbs. However the throbbing of the heart is not why it was selected.

Sleep clearly has functions that are selected for. During the night – when in our original ecological niche there is nothing much to do for humans – it makes sense to replace and repair old or damaged cells and tissue, to save energy and to restore the levels of chemicals we need during the day. This indeed happens while we sleep.

During NREM sleep pituitary growth hormones peak. These hormones are involved in the growth of children and in the making of new cells and repair of damaged tissue in both children and adults. Also, the metabolic rate is slowed down, which conserves energy and lowers the amount of food that is needed. While we sleep the levels of the chemicals we used during the day are also restored. Two of these chemicals, as we, as we already saw, are serotonin and epinephrine which are needed during the day for attention, learning en memory.

Dreaming does not seem to be of any assistance to those functions. It is more likely that our brains are evolved – amongst other things – to make sense of our experiences while awake: “The mind-brain’s day job, one of its important jobs at any rate, is to make sense out of experience so that we can negotiate the world successfully.” (Flanagan 2000: 140) During our sleep this system simply is not turned off, so it tries to make sense of our experiences. This also is the position of Rodolfo Llinás and Denis Paré (1991). They argue that the difference between the sleeping and the waking state is that in the waking state *sensory input* is processed by the thalamocortical system.
However, this processing of sensory input is just a small part of the activity of this system and there is no further difference between the activity during waking and sleeping. This leads them to claim “that wakefulness is nothing other than a dreamlike state modulated by the constraints produced by specific sensory inputs.” (1991: 525) The fact that the processing of sensory input is just a small part of the activity of the thalamocortical system also means that the “the thalamocortical network appears to be a complex machine largely devoted to generating an internal representation of reality that may operate in the presence or absence of sensory input.” (1991: 525). This places the phenomenal experiences firmly inside the head.

At least in the case of dreams phenomenal presentations are internal and interpreted as representations. Even if the phenomenal experiences are not representations – but mere presentations – the brain will still interpret the experiences as being representational and try to come up with the best story about what happens, which in the case of dreaming is the least incoherent story. When we are awake there are sensory inputs, which place restrictions on the interpretation, for there will be a constant flow of information about the situation the subject is in. In dreams one can fly, for there is no sensory input that contradicts this interpretation of the presentations. In normal life one will not have the sensation that one is flying, for the sensory input shows that we are for instance walking in the street.

The conscious mind then functions as an interpretation device that constructs a representation out of the experiences in a way that fits the world best. If the presentations are indeed also representations, the interpretation is restricted by the properties of the world.

The thesis that the phenomenal mind is an interpretation device that is selected to generate a representation of the environment that is as coherent as possible is also supported by the phenomena of the Charles Bonnet syndrome. The Charles Bonnet syndrome is a condition in which patients have vivid hallucinations in combination with normal visual perceptions of the environment. In general,

(1) the visions occur in a state of clear consciousness and are not accompanied by another disorder of sensory-deception; (2) they coexist with normal visual perceptions and the individual retains the ability to differentiate them from the latter; (3) the visions appear suddenly and unexpectedly and tend to disappear on closing the eyes; (4) they are not accompanied by any particular emotional reaction. (Damas-Mora et al. 1982: 252)

Part of the cause of Charles Bonnet syndrome is bilateral reduction of the visual input (White 1980; Teunisse et al. 1995), mostly due to a bilateral cataract. The brain receives some input from
either the retina or another part of the visual system and interprets this input. But since the input is unspecific, no adjustment takes place, when the interpretation is incorrect, since the minimal input is coherent with the representation that is generated. Geoffrey Schultz and Ronald Melzack (1991) argue that the hallucinations of the Charles Bonnet syndrome should be explained as follows.

The complex visual experiences of the Charles Bonnet syndrome are generated by neural-impulse patterns, originating in those brain areas subserving vision. These patterns of neural impulses may be initiated by nonspecific input of the ascending reticular formation, activity of intact visual receptors, or hyperactivity of cells central to the damaged visual areas. This theory emphasizes the inherent capacity of the brain to generate, or construct, meaningful experience without direct correspondence to sensory stimuli. (1991: 823)

In sum, dreams show that we have a natural inclination to interpret presentations as representations, a claim that is supported by the Charles Bonnet Syndrome. Furthermore, this interpretation device is not turned off during sleep, and might even be active when we try to turn it off deliberately as in the case of general anesthesia. Dreaming also places the phenomenal experiences and their content firmly inside the head, thereby showing that internalism about phenomenal content has to be right.

This latter suggestion also followed from the application of the natural method to synesthesia. I will now show that in the case of colors a secondary property view has to be right.
CHAPTER TEN
CASE STUDY # 4: COLOR EXPERIENCES

In chapter eight the question was raised whether in normal situations phenomenal color experiences represent the world as it is or not. Color realism is the view that colors exist independently of us: fire trucks and ripe tomatoes are really red. In this chapter I will argue in favor of color anti-realism: there are no colors independently of us. If there are no organisms that experience the world like we do, fire trucks and tomatoes are no longer red. Our experiences of colors are caused by primary properties of the objects we believe to be colored. In other words, there are no colors, only color experiences and their causes. Talk about color, though, is very convenient, for things indeed seem to be as color realism tells us they are. I will use the natural method to show that colors are secondary properties. In this case the natural method consists in combining phenomenology with the psychology and physics of colors.

We can argue that something is a secondary property in two ways. The first one is to show that the property has to be dependent upon the observer. Secondly, one can argue that it is impossible to identify the property in the absence of the observer. I will use both strategies, respectively in sections three and four of this chapter. I will start by saying something about my own phenomenology of color experiences and present John Locke’s description of primary and secondary properties which I think still holds.

1. the phenomenology of color experiences

For most people the phenomenology of color experiences will be pretty straightforward: ripe tomatoes are red, the sky is blue, and bananas are yellow. However, things are not so easy for people like me who are colorblind (I am a deuteranope). Contrary to what noncolorblind people often think, there are very few people that experience only black and white: most colorblinds experience many colors, but mix them. For instances, when my brother, who is not color blind, drives me home from a visit to my parents we have to take a road full of traffic lights. This road is lit with – as he says – yellow neon lights. The problem for me is that I do not experience any difference between the yellow lanterns, the red traffic lights, and the taillights of cars in front of us: the road is a sea of red lights to me, which would cause problems if I were driving, for I might take a red traffic light for a yellow lantern. However, I do experience lemons as yellow and tomatoes as red. This is typical for the most common form of color-blindness, called
deuteranopia or green-blindness, where the eye lacks the cones that are sensitive to the middle wavelength. Though colors like reds and greens are confused, all colors are experienced (McIntyre 2002: 36).

Since it is clear that there is something wrong with the cones in my eyes, this seems to indicate that I structurally misrepresent the properties of the lanterns. The property yellow is represented as red. According to the criteria for misrepresentation described earlier, this would indeed be a misrepresentation, for (1) there really is a property of a object that is represented; and (2) this property is not presented as A as it should be, but as B. However, for me my deviant color experiences always have been a strong basis for the intuition that colors are dependent upon humans, and not only upon the properties of objects that are still present in the absence of humans. This does not mean that I do not misrepresent a property of the lantern, but if it is a misrepresentation, it would be a very special one, since I could not have detected it without help from noncolorblinds (Cf. the example of the rat that was misrepresented as a cat in chapter eight: this kind of misrepresentations can be detected without the help of conspecifics). In other words: colorblinds do not have any trouble in their normal life with their interaction with the natural world, only with their noncolorblind conspecifics. It is only according to the latter that someone is colorblind. What if everyone had the same cones as I do? What would happen to the color of the lanterns? Intuitively I would say that we all would agree that they are red (hence should be replaced as soon as possible with white lights).

I will return to this “planet of the colorblind” scenario shortly, for a version of it is not as fictional as it might seem at first sight. First I will say something more about the discussion about primary and secondary properties.

2. John Locke on primary and secondary qualities

Traditionally, it is John Locke (1690) who made the distinction between primary and secondary properties famous. In making this distinction he followed his mentor Robert Boyle. At the end of chapter eight of the second book of An Essay Concerning Human Understanding, Locke describes the three different sorts of qualities bodies do have. I believe these descriptions are still entirely correct.

The qualities then that are in bodies rightly considered, are of three sorts:
First, the bulk, figure, number, situation, and motion or rest of their solid parts. Those are in them, whether we perceive them or no; and when they are of that size that we can discover them,
we have by these an idea of the thing as it is in itself, as is plain in artificial things. These I call primary qualities.

Secondly, the power that is in a body, by reason of its insensible primary qualities, to operate after a peculiar manner on any of our senses, and thereby produce in us the different ideas of several colours, sounds, smells, tastes, etc. These are usually called sensible qualities.

Thirdly, the power that is in any body, by reason of the particular constitution of its primary qualities, to make such a change in the bulk, figure, texture, and motion of another body, as to make it operate on our senses differently from what it did before. Thus the sun has the power to make wax white, and fire, to make lead fluid. These are usually called powers. (section 23)

The debate on whether Locke was right is still going on. The program for the rest of this chapter is the following. In the next section I will discuss the scenario in which we supposedly are all color-blind. I will do this in order to show that colors are indeed secondary properties without thinking about what the properties out there might be that we call “red,” “yellow,” and “blue.” After that I will present Frank Jackson’s primary view of colors and will discuss some data collected by Edwin Land that show that this primary view of colors cannot be right, for there are no properties out there that we can call “red,” “yellow,” and “blue.”

3. the planet of the colorblind

At the beginning of this chapter I asked the question what we would think about colors and color realism if everyone had my color experiences. In the planet of the colorblind scenario everyone would be colorblind. I believe that a version of this scenario might actually be true, so this is not meant as the starting point of a conceivability argument, which – as I argued in chapter two – will never provide us with new knowledge.

In this section I will start by saying something about the neurology of color experiences, which makes it easier to explain the different kinds of color vision among humans. I will give an overview of the differences, where I explain how the neurology of colorblind people differs from those with normal color vision. After this, I will turn to the genetics of anomalous trichromatic color vision, which suggests that most humans with normal color vision can be considered to be colorblind too. Some evidence for this latter claim will be presented.

This leads to the conclusion that a version of the scenario of the beginning of this chapter is not that farfetched after all.

3.1 the neurology of color vision

Even though signals are processed in a parallel manner, the brain also works in a serial way. If we take a look at color vision, this serial process starts at the retina. The retina consists of several
layers, of which the first contains of two kinds of photoreceptors: rods and cones. In subjects with normal color vision the rods and cones in the first layer of the retina are stimulated by the different kinds of light falling on the retina. The cones are located in the center of the retina, the fovea. Cones are usually thought to come in three kinds. “There are three types of cones[...] Their collective sensitivity covers the range of the electromagnetic spectrum from a little less than 400 nm to a little more than 700 nm.” (Hardin 1993: 26) The three types of cones are named after the wavelengths they are sensitive to. The first type is the short-wave-cone, which is maximally sensitive at 426 nm (blue). The second and third are the middle wave-cones and the long wave-cones, respectively maximally sensitive at 530 nm (green) and 560 nm (red). Normally the fovea is used to see things in detail. The rods are located in the periphery of the retina.

Below this layer of photoreceptors is a layer of ganglion cells of which the axons form the optic nerve, sending the signals from the rods and cones to the layers of the lateral geniculate nucleus (LGN). The LGN has six layers of cells: the upper four layers are the parvocellular or P layers and they consist of small cells that have some wavelength specificity. The other two layers are the magnocellular or M layers and consists of lager cells that have no wavelength specificity (Zeki 1993: 180).

From the LGN the signals go further to the primary visual cortex, also known as area V1 or the striate cortex, which lies in the occipital cortex. Lesions in this area lead to cortical blindness, either completely or in one part of the visual field. There are four routes from the LGN to different layers in V1.

First, there is the pathway that starts in the M ganglion cells of the retina to the M layers in the LGN and from there to layer 4B in the visual cortex. From there the signals travel further to area V5. Lesions in this area show that this is the part of the prestriate cortex where visually perceived motion is processed. A second pathway travels has a similar route from the M ganglion cells in the retina to the M layers in the LGN to layer 4B of V1. After this the signals are relayed to area V3 of the prestriate cortex, which is a form processing area. The ether two pathways are pathways that have their origin in the P ganglion cells of the retina and go to the P layers of the LGN. From there the signals are send to layers 2 and 3 of the visual cortex. The route from these layers further goes in both cases to area V4, which contains cells that are form-selective and cells that are wavelength specific. So, here color and form are processed in the same area (Zeki 1993).
3.2 cerebral colorblindness
Colorblindness comes in two kinds: cerebral and retinal. Most cases of cerebral achromatopsia are cases in which there is damage in area V4 (Zeki 1993: 268). Of course, when the pathway to V4 is damaged, subjects cannot experience colors either. This was probably the case with Jonathan I. A painter who became an achromat after he was hit by a truck. Oliver Sacks (1995: 31) suggested that this man had damage to the pathways from V1 to V4, because the CAT and MRI-scans of this area were normal. Another explanation might be that these scans were not accurate and that a PET scan would have shown that it indeed was V4 that was damaged. I. Died before this technology was available. Interestingly, this man was a synesthete who experienced music as colored prior to his accident. After the accident he also lost his synesthetic experiences, as well as colored dreams and colored hallucinations during migraine.

Most cases of colorblindness, however, are not cerebral but retinal. The varieties of colorblindness discussed in the following section are all varieties of retinal colorblindness.

3.3 retinal colorblindness
Retinal colorblindness roughly comes in two kinds: total and partial colorblindness. Total colorblindness, evidently, comes without varieties. Partial colorblindness comes with different varieties, and even in normal color vision there are anomalous cases that are closely related to the varieties of partial colorblindness.

Achromats /monochromats
Less than 1 in 33,000 people are achromats. These individuals have no or no working cones at all. This results in having no color experiences. Since the cones are located in the fovea, they also are unable to see things in detail without magnifying glasses. Furthermore, they are extremely sensitive to light, which means that they have to wear sunglasses every time they leave their houses.

 Dichromats
The most commonly found varieties of colorblindness belong to the category of dichromacy: 1 in every 20 white males has this kind of colorblindness (it is exceptionally rare in females or noncaucasian males). There are three varieties depending on which cone is missing (or not functioning). Those who are missing the long wavelength sensitive cones are called protanopes.
Those who are missing the middle wavelength sensitive cones are deuteranopes and those who are missing the short wavelength sensitive cones are called tritanopes.

In addition to the problem of color discrimination protanopes also have a reduction in brightness as to the point that some colors (e.g. red) might be experienced as gray or black (hence a traffic light might seem not to work). Protanopes have trouble making distinctions between red, orange, yellow and green, or between violet and blue. Phenomenally, protanopes and deuteranopes are very similar. Both groups of individuals know that they do have a color deficiency and they have trouble distinguishing the same kinds of colors. Together protanopia and deuteranopia are also known as red-green colorblindness.

**Trichromats**

Contrary to dichromats, trichromats are able to spot ripe (red) fruit at once. Many animals that eat fruit though are colorblinds: “Man and apes and monkeys of the Old World are the only mammals with tree-colour vision.” (McIntyre 2002: 33) The other mammals lack the cones sensitive to the long wavelength (“red”), hence they cannot distinguish between reds and greens, and will have the same trouble as protanopes and deuteranopes spotting ripe fruit among the green of the environment.

**Anomalous color vision**

Besides missing cones or having cones that do not work, there are also milder forms of color deficiency, known as color anomalies. Protanomaly is also known as red-weakness. In the case red, orange, yellow, yellow-green, and green, are somewhat shifted in hue towards green. Even though a protanomalous individual is a trichromat, this might result in a well-known phenomenon, viz. that of a discussion about the color of an object that lies close to another color for all trichromats (e.g. is the book cover blue or purple?). For a normal trichromat there is some red in an object that is said to be purple. For the protanomalous observer, however, the red is shifted, and might therefore just be missing in the color of the object. This means that to him or her the color of the object seems to be blue. Protanomalous observers also have a loss in brightness of colors compared to normal trichromats.

Deuteranomaly, or green-weakness, is similar to protanomaly, only in this case red, orange, yellow, yellow-green and green are somewhat shifted towards red. Contrary to the former case, deuteranomalous observers do not have a loss in brightness. Protanomalous or deuteranomalous
individuals mostly do not encounter any problems in everyday life, though they perform very
differently from normal trichromats on color vision tests. Most of the time they do not even
know that they have a color deficiency.

3.4 the genetics of protan and deutan color deficiency

Combining the phenomenology and neurology of color vision and color blindness with the
genetics of protan and deutan color deficiency shows that most people are in fact color blind,
hence that the planet of the colorblind scenario is not just a conceivable scenario but an actual
one. If we take a look at the genetics of anomalous trichromacy, then we get the following
picture. If a son has a protan or a deutan color deficiency, then the mother will be a carrier of this
deficiency, since the relevant chromosome is the X chromosome. The genes responsible for the
development of the long wave and middle wave sensitive cones are located of the X-
chromosome. Normally the deficient gene is recessive, meaning that a woman having both a
normal and a deficient gene has normal color vision, and a man with the deficient gene will
always be color-blind. If a son is an anomalous trichromat, then the mother is heterozygous for
the color deficiency: she has two X chromosomes, of which one is responsible for deviant cones
to develop, and the other for normal cones to develop. (The story for tritan color deficiency is
different, for the gene is located on the seventh chromosome, which is an autosome.)

However, sometimes the dominance of the normal chromosome over the recessive one is
cancelled. In a fertilized egg that contains two X chromosomes both chromosomes are active.
Early in the embryogenesis one of these X chromosomes becomes inactive in each cell. This is a
random process, that is called X chromosome inactivation. If a cell divides then the persecutors
contain the same active X chromosome.

On average approximately one-half of the cells have a specific active chromosome. But chance
might favor one or the other X chromosome in a particular tissue. For this reason females who are
heterozygous for an X-linked recessive trait show some expression of the trait. At the cellular level
each cell is found definitely to express one or the other allele, but not the other. (Sutton 1997: 609)

This seems to suggest that mothers of men with a particular color deficiency express this
deficiency to some extend themselves. But there is also another suggestion, and this is the one I
am interested in, for it might show that the planet of the colorblind scenario is not that fictional
after all.
The suggestion is the following. There should be some women which have four different cones: tetrachromats. If, in addition, their brains would be able to make use of the fourth cone, then they should have even more color experiences than normal trichromats. This would imply that they would be living on a planet inhabited with conspecifics they consider to be colorblinds.

This scenario might seem somewhat speculative: can there be a tetrachromat, or is this just a logical possibility but not a physical one? There are two main reasons to think this is a real physical possibility.

First, there are animals that have four or more cones. Many birds do have tetrachromatic vision, and so do goldfish. There is even a variety of shrimp that has ten different kinds of color receptor, though as McIntyre notes, it is not clear what this means to the shrimp (2002: 29). The same question remains for a species of desert ants that has four types of photoreceptors and a butterfly that has five (2002: 30).

Secondly, it is already known that all males of the squirrel monkey are dichromats. Two thirds of the females, however, are trichromats. These females are able to make discriminations in the red-green range that their dichromatic conspecifics cannot. It is precisely due to X chromosome inactivation that they gain an extra cone, and apparently they do have brains that can make use of these extra cones (Mollon et al. 1984; Jacobs 1984; Toveé et al. 1992).

So, tetrachromats do exist on our planet and X-chromosome inactivation is able to produce trichromacy in females of an otherwise dichromatic species. Why then should X chromosome inactivity not be able to create human female tetrachromats?

3.5 Mrs. Tetrachromat

Gabriele Jordan and John Mollon (1993) asked how they could find a tetrachromat. Their starting point was to look for mothers of men with anomalous trichromacy, since these men do have three cones, but the peak intensity of one of them is slightly shifted to either the green (protanomaly) of the red (deutanomaly). This means that their mothers might have four cones: the three normal cones and one of the slightly changed cones. X chromosome inactivation can account for having both the deviant cone and its normal variety.
Now, in addition to having four cones a *strong tetrachromat* should also have the brain structure to actually use the extra cones. If not, then these women are called *weak tetrachromats*, and it has been determined that they indeed exist (Nagy *et al.*, 1981). Since most men with a color deficiency are either protanomalous or deuteranomalous, Jordan and Mollon used a test in which the short-wave cones become useless, by using bundles of light in the middle- and short-wave range. Normal trichromats become dichromats for this test, for there is no wavelength that excites the S-cones.

If a trichromat looks at a chromaticity diagram then there are three corners of it that are blue, green and red. All other colors are in between. For a dichromat one can draw confusion lines or pseudo-isochromatic lines in this diagram (McIntyre 2002: 42). All points on this line have the same color to a dichromat. These point all look different to a color normal, for a normal trichromat will be able to see the part of the spectrum the dichromat is blind to. (Consider the following analogy: in a normal three dimensional diagram one cannot distinguish whether a line is coming towards one, or moving away if one has only the X- and Y-coordinates. All lines with the same X and Y-coordinates appear the same, even though they might be different to someone who also knows the Z-coordinate.) This fact is used in the Raleigh test for color blindness. For a protan or a deutan any mix of red and green light can be matched to a monochromatic yellow light (providing the brightness of the yellow light may be adjusted). Color normals have just one match (and the brightness does not matter).

So, using a mix of green and red light normal trichromats can create a colored light that is phenomenally similar to a monochromatic yellow light. If normal trichromats are asked to make a match for – say – an orange light, they can make create multiple matches using mixtures of either green and orange or yellow and red lights. The prediction was that a strong tetrachromat would have an extra cone that has its peak sensitivity between green and red, hence they become trichromats in this experiment and should be able to distinguish between the various mixes of light, resulting in finding just one combination for the target light. (By analogy: the trichromats only have access to the X- and Y-coordinate, while the tetrachromat also has access to the Z-coordinate.) Of the 31 carriers of either protan or deutan color-deficiency one women indeed showed the predicted results. Jordan and Mollon are careful and do say that this does not conclusively prove that this women is a tetrachromat, but the result strongly suggest that she is. What adds to the evidence, is that the woman herself reports that her color vision is different,
that “People will think that things match, but I can see they don’t.” (Mrs. M. quoted in Zorpette 2000).

3.6 conclusion

Tetrachromats then, may well exist and would have more color experiences than normal trichromats. To them normal trichromacy is just a mild variety of colorblindness. If we would be color realists, this would mean that most humans are colorblind and are misrepresenting the colors objects actually have. The evidence presented in this chapter shows that this is not the case: colors of objects depend both on the primary properties of the surface of the object and on the observer. Colors are experienced and immediately projected back onto the cause of the experience. That is why the tomato seems to be red even though it actually is not. Since we do not know what the primary properties of the tomato (and its surroundings) are that cause the experience of red in us and it is a phenomenal fact that it is the tomato that seems to be red, it is a useful matter of speech to say that the tomato is red. This then is not a misrepresentation: as far as we can use the manner of speech, this is what we should say.

Having shown that colors are dependent both on the primary properties of an object and on the observer, I have shown that colors are secondary properties of these objects. As I said at the beginning of this chapter, there is also a second way to argue for the secondary property view of colors: we can show that there is no (type of) property in the world that can be identified with a (type of) color. E.g. there is not one property that all red objects do have in common. This is what I will do in section five. First I will present the primary quality view of color in section four.

4. the Australian view

Frank Jackson (1996) defends “the ‘Australian’ view that colors are physical properties of objects.” (1996: 199) He describes colors as primary properties: “[Colors] are identical with complexes of the properties the physical sciences appeal to in their causal explanations.” (1996: 199) In chapter two I used this criterion to mark the difference between primary and secondary properties: only primary properties can function in the causal scientific laws.13 So, I agree with Jackson, that if colors are primary properties, they should be identified with what I believe to be the external cause of the colors. Jackson’s defense of the primary quality view of color depends upon our common sense view of color: the grass really is green and fire trucks really are red. He calls the intuition that objects are colored the prime intuition (1996: 200).

13 I take it that experiences are primary properties of experiencing beings. These experiences are projected onto the object that caused them, creating secondary properties.
Jackson discusses a problem for this view: looking close at a surface that is red from a distance, one may see that the surface is build up from small spots of yellow and magenta. Can the primary quality view of color determine what the real color is? Yes, it can. Jackson does this by introducing viewing conditions.

What we need to say [...] is that the color something has in [...] one circumstance may differ from the color it has in another, where the viewing distance is part of the circumstances and that each color is equally 'real.' (1996: 205)

I think this is the right thing to say for a proponent of the primary quality view of color. Even though primary qualities are independent of an observer, it depends on the level at which an observer is looking which primary qualities can be observed. Consider the primary quality of *shape*. If a circle is made up of sugarcubes, then the viewing distance determines whether one sees a circle, or a cube. This just is an implication of the mereological model of the world, and it implies that all the properties of each level are equally real. If colors are primary properties, then this would also be the case for colors. This then is not a problem for the primary quality view of color, and Jackson rightly defends this as a part of that view.

4.1 the psychology & physics of color experiences

Arguing against the view that colors are out there, we should consider the following two options. Either colors are properties of light or of (surfaces of) objects. In either case, there are two further options: either there is one property that objects or light of a certain color have in common, or this color is a disjunctive property. If it is a disjunctive property, then it should not be excessively disjunctive, for then we would be talking about different *kinds* of properties. This analysis of the different options for the primary quality view of color provides us with a strategy to show that it is wrong. The first step is to show that both in the case of light and of objects the possible candidates for identification with color results in the acceptance of excessively disjunctive properties. This implies that the hypothesis that each color can be identified with either one property of an object or one property of light is refuted as well. The conclusion would then be that the proponent of the primary quality view of color has to accept different kinds of – say – redness. Step two will be to show that this is not a defensible position. Before applying this strategy to show that the primary quality view of color is false, there is a question that should be addressed first: *when are disjunctive properties excessively disjunctive?*
Jackson agrees that a disjunctive property should not be excessively disjunctive, for then we would have not one disjunctive property, but of two (or) more separate properties:

Consider, for instance, the sentence 'Either arsenic administered by Harry or cyanide administered by Mary caused the death'. Surely we only make sense of this sentence by reading it as 'Either arsenic administered by Harry caused the death or cyanide administered by Mary caused the death'. (Jackson 1996: 212–213)

What might be the nonexcessively disjunctive properties that can be identified with the different colors? Jackson himself seems to think that wavelengths of light are the best candidate:

Semir Zeki (1993: 228) notices a paradox in the common sense and primary quality view of color. On the one hand the intuition is to identify colors with wavelengths, but on the other hand there is color constancy.

And here enters the paradox. When we view objects in different conditions of illumination, the wavelength composition of the light reflected from them changes. If, for example, one were to view an orange or a banana in a room lit by tungsten light, and then in a room lit by fluorescent light and then, successively, in daylight on a cloudy day and on a sunny day, and at dawn and at dusk, one would find that the orange will continue to look orange in colour and the banana will continue to look yellow. There may be some changes in the shade of yellow and orange, but the colour will remain the same. Yet, if one were to measure the wavelength composition of the light reflected from these surfaces in these conditions, one will find profound variations. In natural viewing conditions there is thus no specified wavelength composition, or code, that leads to a particular colour and to that colour alone. (Zeki 1993: 228)

Though I believe this shows that the view that colors are identical to wavelengths is excessively disjunctive, these different triplets of wavelengths are what Jackson seems to accept as a proper disjunctive properties.

**Excessively disjunctive property of light**

If this argument from color constancy is not enough to reject the primary quality view of color, I believe that there is also another way to show that there can be no such thing as a disjunctive property of blueness (or any other color). The argument has the following structure. The property has to be a property of an object or of light. If we can show that such a property can be part of both blueness and redness — without altering the viewing distance — then the primary
quality view of color has a problem, for the disjunctive property contains too much: it might not be excessively disjunctive to take \( X \) and \( Y \) (where \( X \neq Y \)) as two parts of one disjunctive property, but it is clearly incoherent to take \( X \) also to be a part of a totally different property, for that would nullify the distinction between – say – blueness and redness. This then provides us with the direction in which we should be looking to disprove the primary quality view of colors: if Jackson’s suggestion that triplets of light are colors then the same triplets are not allowed to identify different colors, for then blue would be red.

This is exactly what an experiment by Edwin Land shows. Land made a display which looked like a painting of Mondrian: rectangular pieces of paper of different colors and of different sizes were placed in the display. Such a display is referred to as a “color Mondrian”. Three projectors are used to illuminate the display, projecting long-wave, middle-wave and short-wave light.

A normal subject looks at this display and is asked which color a certain patch has. The display is illuminated in such way that it reflects 60 units of long-wave, 30 units of middle-wave and 10 units of short-wave light. This can be done by using a telephotometer: first only the long-wave projector is switched on, and the intensity is varied until the telephotometer says that 60 units of light are being reflected by the patch. The same is done with the middle-wave and short-wave projector, after which they are all three switched on at the same time. The subject reports a designated patch to be green. The experiment is repeated with other patches, each made to reflect 60 units of long-wave, 30 units of middle-wave and 10 units of short-wave light. By doing this, the triplets of light reflected by each patch are the same. Subjects report that the patches are blue, white, red, etcetera.

This is a serious problem for the primary quality view of color. Each patch was made to reflect the same triplet of light, but the phenomenology of the color-vision did change with each patch. The conclusion that follows is that “when a surface forms part of a complex, multicoloured, scene there is no simple and obvious relationship between the wavelength composition of the light reflected from that surface and its colour” (Zeki 1993: 231). So, if triplets of light are to be identified with color, then the same triplet is part of the disjunctive property redness, greenness, blueness, yellowness etcetera. This shows that colors cannot be identified with triplets of light, for if that would be so, then we could not make any difference between colors.

**Excessively disjunctive property of objects**
If one still tries to defend the primary quality view of color, then there has to be another primary property that can be identified with color. It seems obvious to look at the properties of the surface of – in this experiment – the patches in the color Mondrian itself. However, an extension of the Land color Mondrian experiment shows that this is no option either. If we do not change the illumination of one of the patches of the former experiment, but look at this patch in isolation (the void condition) the phenomenology changes: instead of experiencing green, subjects experience white or very light gray. If one brings the surround into view again, the patch immediately is experienced as green. It does not matter where the green patch is placed on the display: varying the surround has no influence on the experienced color. Each time the surround is illuminated, the reported experience is that of greenness. The experiment can be repeated with other colors, which all lead to the same results. These experiments show that: “the colour of a surface is determined not only by the wavelength composition of the light reflected from it, but also by that from surrounding surfaces.” (Zeki 1993: 232) So, if the surrounding surfaces are causally effective in creating the color, there cannot be an intrinsic property of the surface of the designated patch that is colored, for if that would be the case, the surround would not make any difference.

In sum, we cannot identify colors with either triplets of light, nor with specific surface properties, for the same triplets of light might induce different color experiences in us, and the same surface properties might induce different color experiences in us even when the surface properties of the patch, the light source and the observer are kept constant. Trying to identify color with a property of light or with a property of surfaces leads in both cases to accept that property A and property B belong to the disjunctive property of both – say – red and green. This lead to two problems. First of all, at the level of the alleged primary properties we are no longer able to distinguish between red and green. Secondly, the triplets of light or the surface properties that constitute the disjunctive property of red are so different that it is hard to see how this can be regarded as one disjunctive property instead of as a disjunction of totally different properties, for instance the surface of a fire-truck has totally different properties than that of a tomato. This means that the primary quality view of color needs to accept more than one kind of – say – redness. This is what Jackson does: “Indeed, I think that we could live with considerably more than two rednesses.” (Jackson 1996: 216)

I believe that this is no solution, for the problem for then we have to distinguish between what is red, and what is not red but only looks red; ‘fools’ red, as Jackson himself calls it. In the case that there are more than two rednesses, there would be several kinds of ‘fools’ redness and only one
kind of redness. The case is analog to the H2O-XYZ case: XYZ is not water, for water is H2O. If we say that K is the only kind of redness, then all other kinds have to be kinds of 'fools' redness, which will be the huge majority. But if something is not red, but only looks red (and does so all the time, say ripe tomatoes are 'foolishly' red), then what is it that causes the experience of red in us if it is not the redness of the tomato? Surely, we don't experience a difference between the redness of a fire truck and that of a tomato: both are the same kind of experience. But, this would imply that the red experience caused by a tomato is caused by some other property of the tomato than its being red, hence by another primary property. In other words the 'foolish' redness of the tomato has to be a secondary property. And since most kinds of redness are fools red, this will be the case for most of them. That would mean that most experiences of red are caused by things that are not red at all. This, evidently, is contradictory to the primary quality view of color, which then has to be false.

4.2 conclusion

So, starting both from experience and from the properties of light and objects, it can be demonstrated that colors cannot be primary properties and have to be secondary properties. But this leads to a further question: what good is it to think objects are colored? We have representations of primary properties that are indirect representations: e.g. the experience of red is used as an indirect representation of the primary properties of a ripe tomato, and we do not know what these properties are. In the case of primary properties and a direct representation the way of representing makes good sense: a hard object is represented as hard, and a square is represented as a square. But why should we use phenomenal representations and project them back as if they were properties of the object? This seems to come very close to misrepresentation, for A is represented as B and not as A. What good is it to think that a tomato is red? Why use primary properties of surfaces to create experiences, which in turn are projected back on the cause of the experiences, leaving the representing creature under the illusion that the object does actually have these properties? This is a question I will answer in the next chapter.
CHAPTER ELEVEN

CASE STUDY # 5: EXPERIENCES OF BEAUTY EVOked BY FACES

Earlier I claimed that phenomenal consciousness is a heterogeneous set of (re)presentational states, which are either selected for in evolution, or are spandrels such as dreams and afterimages. In this chapter I will focus on the evolutionary aspect of phenomenal consciousness and will demonstrate that they are evolution’s shorthand by applying the natural method to experiences of beauty evoked by faces. Why do we experiences faces of other humans as being beautiful? Does beauty lie in the eye of the beholder? And if so, is the experience of beauty a matter of subjective preferences, or are there people that have faces everybody finds beautiful, regardless of sex, sexual preference, race or culture? If there are certain face-types everyone experiences as attractive, can we explain this?

Just as in the four other case studies I will use the natural method to search for answers to the problem of phenomenal consciousness. In this case this means combining phenomenology, psychology and biology. The data I will present in this chapter show that beauty does indeed lie in the eye of the beholder, but that experiences of beauty are also intersubjective. Experiences of beauty come in different varieties, depending on what it is that evokes these experiences. We can experience beauty when we look at natural phenomena such as other people or landscapes. But we also experience beauty when we perceive cultural phenomena, such as music, paintings or poetry. I will mainly focus on the first example in the natural category, for in this case we can be sure that the represented objects were part of the environment in which we evolved. This means that mostly the feeling of beauty is also a feeling of attractiveness, which is not, of course, the case if we would be looking at landscapes or art.

1. **the phenomenology of experiences of beauty evoked by faces**

Being beautiful is something that is highly appreciated in our society as well as in others. We want to have partners that are beautiful, and some individuals really have become *models* for others: persons requesting plastic surgery often want to have the physical features of famous people. In the United States of America, where over two million people have cosmetic surgery each year (Haiken 2000: 83), the models that are requested the most are movie stars like Winona Ryder, Johnny Depp (eyes), Sandra Bullock, Harrison Ford (jawline), Salma Hayek and Mel Gibson (body) (Simpson 2001: 3). Some famous people – like Michael Jackson – go through considerable length themselves to change their facial appearance. Subjects do have very specific demands
when they want to change their appearances. For example, the performance artist Orlan wanted plastic surgery so she would end up having a face with “the chin of Botticelli’s Venus, the forehead of the Mona Lisa, the eyes of Gérome Psyche, the mouth of Boucher’s Europa, and the nose of a Diana from the school of Fontainebleau.” (Bruce & Young 1998: 127-129) Even though not everyone wants to look like Harrison Ford or Sandra Bullock, nobody will have plastic surgery to have the eyes of Marty Feldman or the chin of John Goodman. This indicates that there might be an intersubjective standard of facial beauty. If this is true, then most people are not aware of it and even think this claim is false, since it belongs to the realm of general belief that beauty lies in the eye of the beholder and that one’s opinion of what and who is beautiful is largely culturally determined. In the above case the requests for cosmetic surgery would be determined by Hollywood.

If there would be no intercultural and intersubjective standards of beauty, then one should expect very different requests for plastic surgery, which is not the case. Can we indeed show that we all judge the same faces to be beautiful, and if so, can we explain why there is some universal standard? If we can do this, what does this tell us about phenomenal consciousness? To answer these questions, we have to apply the natural method and combine phenomenology with the data provided by psychology and (chemical) biology.

2. the psychology of experiences of beauty evoked by faces

Victor Johnston has done research into the question which faces are experienced as being beautiful. Being an evolutionary psychologist he is interested in average and ordinary features of humans.

For evolutionary psychologists [...] questions about the average and ordinary are of paramount concern. The average often reflects what is, or was, most adaptive for biological survival, and individual variations are best regarded as potentially important explorations around this successful theme. (1999: 131-132)

Now the interesting question rises: are average faces beautiful? If this is so, then beautiful faces might reflect properties that are beneficial to survival. Judith Langlois and Lori Roggman (1990) demonstrated that the average face indeed is beautiful. They used a software program to create average faces: pictures were scanned into the computer, which averaged the different features of the faces. In this manner images of male and female faces were generated by using 4, 8, 16, or 32 pictures of real faces. Subjects found the faces of both men and women increasingly attractive the more faces were used to create the image. Their conclusion was that average faces indeed are
beautiful (1990: 119). Also Johnston (1999) has used image-generating software to generate average faces. Using many different faces as input, he generated pictures of faces that had average features. These faces were indeed evaluated as being more beautiful than the faces of the real persons that were used to generate the pictures of average faces (1999: 145-146).

Having established both what the average face looks like and that this face is considered to be beautiful, Johnson did another experiment, to see whether these faces were the most attractive. He used a program called FacePrints, which makes use of our recognition abilities. Humans are very good at recognizing people. Recall, on the other hand, is not one of our best capacities. That is why sketch artists often draw pictures of wanted persons that do not look like the criminals the police are after. How does the FacePrints program work? Say you have been witness to a crime and you want to generate a picture of a suspect you saw. This program shows you a face on a computer screen. You can rate it between 0 and 9 depending upon the likeness to the man you saw. The program then shows you another face, and another, and another, and each time you press a key from 0 to 9. The computer program makes use of algorithms that mimic gene transmission, in this case the genes that encode for the different facial features. Each time you evaluate someone not as 0, there is some information about the face that is used in one of the next faces you see. “The fittest faces would be given a higher rating, and their “children” would inherit their genes.” (1999: 43) In the end a picture is generated that is rated 9, which indeed looks more like the wanted suspect, than a sketch based on recall (1999: 50).

Johnston used the FacePrints program to generate beautiful female Caucasian faces. The ratings 0 to 9 now were used to indicate the beauty of the face. The faces that were created were not average at all. Men and women generated faces with very specific features that significantly differed from the features of the average female face (1999: 146). What are these differences? The two most profound differences are that for the more attractive females the lower jaw was smaller than in the average face and that – while the width of the mouth is smaller – the lips are fuller-than-average. Other differences are that the eye-hair distance is larger, while the eye-chin, the eye-nose, the eye-mouth and the mouth-chin distances are smaller in the more beautiful faces than in the average faces (Johnston & Franklin 1993: 190-191).

**Two concerns about these data**

One might be concerned about the intersubjectivity of these findings: is beauty not *culturally* determined? Is it not the commercial industry and Hollywood that tell us who we should find
beautiful? The requests for plastic surgery surely seem to suggest so. This was a serious problem for those who searched for universal standards of beauty, but it has been proven wrong. First of all, Johnston’s experiment has been replicated in other cultures with the same results (Perrett et al. 1998: 884-885; Johnston 1999: 146). Furthermore, two related studies yield the same results.

The first of these related studies concerns experiences of beauty evoked by the female body figure. The female waist-to-hip ratio (WHR) that is considered to be the most attractive is 0.7 (Singh 1993; Singh & Young 1995). Again, these results were replicated in other cultures.

Evidence comes from judges of both sexes, different age groups, and different social classes, and it has been replicated in different countries like the USA, the island of Azore, Germany, Great Britain, Guinea-Bissau, Hong Kong, India, Indonesia [...]. It appears that Singh has found a morphological feature which is [...] universally regarded as attractive in women. (Henss 2001: 502-503)

A second study that also shows that (facial) beauty cannot be culturally determined concerns the experiences of infants ranging from three to six months of age. Langlois et al. (1991) have shown that children in this category prefer attractive faces to nonattractive faces. These results were obtained by first selecting pictures of faces that are beautiful and those that are not. The judges, of course, were adults. Subsequently, these pictures were shown to young infants (a picture of both categories at the same time, one left, one right). The looking time at these stimulus pictures was measured, and the result was that young infants looked significantly longer at the attractive faces than at the nonattractive faces (1991: 80-82). These results are interpreted as showing that young infants prefer attractive faces. It is evident that children of this age cannot have been subjected to a lengthy period of cultural indoctrination and have had little or no encounters with media telling them who or what to regard as beautiful. So, the widely held view that it takes exposure to cultural norms and media to learn what is considered as beautiful in the society you live in, is just not true as far as experiences of beauty of other humans are concerned.

A second concern about these results is that the facial features of the most beautiful women might be taken to be all neonate features, and that this might provide an explanation why we judge them as more beautiful. The idea, then, is that we are programmed by evolution to take care of children and it helps if we regard them as being cute. Neonates typically have larger eyes, smaller noses, smaller chins and larger lips. At first sight these seem to be precisely those features of the most attractive female faces (Cunningham 1986: 925), suggesting that this might suggest that the judgment of the most beautiful women is a side effect of our natural tendency to think of little children as being cute. However, Johnston has shown that there are important differences
between youthful and beautiful faces. “Faces that were rated as very beautiful possessed a longer eye-nose distances, slightly smaller eyes, and a wider nose and jaw than youthful faces.” (1999: 156)

In the case of female faces, then, we can conclude the following. Female faces that are judged as the most beautiful do have specific features, such as a smaller-than-average chin and fuller-than-average lips. Though these features are similar to features of young children, they are not exactly the same. These facial features are interculturally judged as being beautiful.

For men, the case is somewhat different, for women judge different kinds of male faces to be the most attractive depending on the point in their menstrual cycle. So, paradoxically there are two kinds of most attractive male faces. If women use hormone-based contraception or are in a low conception risk phase of their cycle, they prefer men with facial features that are a little more masculine than the average male face (which also is experienced as attractive). However, if they are in a high conception risk phase, their preference changes from a face that is just a little more masculine, to a face that has exaggerated masculine traits (Penton-Voak & Perrett 2000: 44). These features include a longer and broader lower jaw, more pronounced cheekbones and brow ridges, and a darker skin than the average male face. Some research has shown that women that are not ovulating prefer a slightly feminized male face, so the results are not consistent here. For my present purpose, however, this does not matter.

3. **the biology of experiences of beauty evoked by faces**

If we take a look at the biology behind experiences of the beauty of faces, we have to consider both the biological features of the person having the beautiful face and those of the person having the experiences. Since both are closely related, I discuss them simultaneously.

During puberty the human body changes. The growth of the body is controlled by androgens – mainly testosterone – and growth hormones. In females the amount of androgens that is released by the adrenaline glands is smaller, which explains why in general women are smaller than men. This not only goes for the length of the body, but also for the lower jaw: during puberty the lower jaws of men grow more than those of women: “so a lower-than-average androgen exposure at puberty would lead to the shorter lower jaw found in attractive female faces.” (Johnston 1999: 148)
A similar explanation can be given for the fuller lips. The fullness of the lips depends upon estrogen-controlled fat deposits during puberty. More estrogen results in higher fat deposits in typical places like the mouth. Full lips, then, indicate a more-than-average estrogen exposure during puberty.

The facial features of the beautiful female face, then, are markers of exposure to low androgen and high estrogen levels during puberty. Since, testosterone inhibits and estrogen stimulates fat deposits in the gluteofemoral region (buttocks and thighs), while testosterone stimulates and estrogen inhibits fat deposits in the abdominal regions (Singh 1993: 294), the same mixture of hormones is responsible for the facial features of the most beautiful women and also for their 0.7 WHR.

So, both the facial features of the more beautiful women and their WHR are makers of the low androgen and high estrogen levels during development. The hormones that determine the bodily and facial features I just discussed, also determine a women's fertility. Very fertile women typically have this mix of high estrogen levels and low androgen levels. So, the markers of this mix are also high-fertility markers (Johnston & Franklin 1993: 183; Perrett et al. 1998: 884).

Furthermore, Singh notes that the 0.7 WHR also indicate health status (Singh 1993: 295). This is corroborated by the evidence that a high WHR (0.85) is correlated with - amongst others - gallbladder disease, heart disease and stroke (Johnston & Franklin 1993: 195).

Can we say a similar thing about the masculine traits? The fact that there is a change in preference when women are in a high conception risk phase suggests that women are attracted to men that have some particularly good features for reproductive success. This indeed is the case. Again, testosterone exposure during puberty plays an important role. As described above, the typical male traits, like the large lower jaw are determined by the testosterone level (which is good for reproductive success as we will see in a moment). Hence, the two kinds of most beautiful male faces (the one slightly more masculine, the other with exaggerated male features) have markers that indicate high and very high testosterone levels.

These data are supported by scientific investigation into male body bilateral symmetry and the preference of women for male odors. The findings are strikingly similar. Women that use hormone-based contraception or are in a low risk conception phase of their cycle have no preference for the scent of males. There are two pheromones in the male scent that play an
important role, androstenol and androstenone, of which the latter is the more prominent odor. Androstenol promotes the female attraction to men, while androstenone does not. In contrast, when women are not ovulating, androstenone induces negative attitudes towards men. When they are ovulating their attitude shifts towards neutral (Grammer 1993). Overall, the evaluation of male scent shifts towards attractive when women are ovulating. Furthermore, women that are in a high conception risk phase do prefer the scent of men that are relatively symmetrical (Thornhill & Gangestad 1999: 191). It is this latter result that supports the results of the facial research, for body symmetry is also a marker of high levels of testosterone. So, attractive men do have certain facial features, are relatively symmetrical and have an attractive scent, which are all markers of high levels of testosterone.

High levels on testosterone are associated with good health, both phenotypically and genotypically, i.e. high levels of testosterone are correlated with “good genes” (Johnston et al. 2001: 262-263). Furthermore, high levels of testosterone are also positively associated with growth rate, longevity, fecundity, and health status (Thornhill & Gangestad 1999: 177; Johnston et al. 2001: 263).

As I noticed earlier, there is some discussion about how to interpret the other most beautiful male face, because results differ on whether this is a slightly masculinized face or a slightly demasculinized face (Penton-Voak & Perrett 2000: 39). Further research will have to be done to determine whether the difference in hormonal status of individual women who are not ovulating influences the preference for a slightly masculinized or a slightly demasculinized face. In either case the evolutionary explanation will shift towards the following account. When women are not ovulating they are not interested in good genes; they are interested in a good provider, a man who will help raise the children. Hence, they are looking for a reliable man, and a man with extremely high levels of testosterone is not a good co-parent. “Indeed, increasing testosterone levels in males is associated with more troubled relationships (including increased infidelity, violence and divorce).” (Perrett et al. 1998: 886)

What we have seen so far is that combining phenomenology, psychology and biology – i.e. applying the natural method – shows that the facial features of men and women that evoke experiences of beauty in us are markers of fertility, good health and good genes. What does this tell us about phenomenal consciousness?
4. experiences of beauty evoked by faces are evolution’s shorthand

Combining the above data with previous conclusions, the following follows. Feelings of beauty, like all other phenomenal experiences are presentations. These presentations can be used as representations. When these feelings are evoked by the physical features of a potential mate, these feelings are nature’s way of telling us that this mate is a good candidate for reproduction, that he or she is very fertile, or in very good phenotypic and genetic health. In other words: experiences of beauty evoked by facial features track fertility, health and good genes. A phenomenal experience that is used as a representation, then, has a double representational content. First, it represents the secondary property of the represented object (the object is beautiful). Secondly, it represents the primary property of the represented object (the object is fertile or in good health). Usually the representing organism interprets the presentation as the former representation and is oblivious to the content of the latter.

Explicitly expressing the nonphenomenal content of phenomenal representations

Humans nowadays are linguistic creatures. This means that we might find out what our indirect phenomenal representations represent. In the case of experiences of beauty evoked by faces we have done so. We now are able to explicitly describe the implicit content of the phenomenal representation, viz. in the form of linguistic statements like: “this man has good genes” or “this woman is very fertile.”

Since it is possible to express the nonphenomenal content of phenomenal representations explicitly, why then did this indirect manner of representing evolve? It seems that indirect representation takes an unnecessary detour. There are two reasons why evolution did not go for the direct representation: (1) direct representation requires more than indirect representation, and (2) even though it seems to make an extra step, indirect representation is much faster than direct representation.14

Direct representation requires more than indirect representation

In order to directly express the knowledge that a conspecific has good reproductive features, the representing creature has to have been developed in such a way that it is capable of language. Being able to develop and understand a language is such a complex ability, that the ancestors of the creatures must have had another way of recognizing the most fertile and healthy partners, either directly or indirectly. This happens when one recognizes a conspecific that is fertile,

14 Again, talk of “reasons”, “intentions” etcetera when discussion evolution has not to be taken teleologically.
because one recognized the conspecific that is beautiful: if one experiences beauty when seeing another creature, one seldom thinks “this person is a good candidate for reproduction.” Even though the recognition is successful, the nonphenomenal content of the representation is not necessarily known. Most of us will recognize beautiful people, while just a small part will believe that these beautiful people are the most fertile people or the ones with the best genes.

Furthermore, the content of the direct linguistic representation is also complex: one has to know at least something about reproduction at the level of transmitting genes before one can understand the sentence “this woman is very fertile.” An information pickup system that has to be endowed with all kinds of complex information before it can find its way in the world surely is something that could not have been evolved without prior evolutionary stages in which the ancestors of the system had to get their information in another way. So, a complex information pickup system can only develop from systems that require less to pick up the information that is relevant for their survival and reproduction.

**Indirect representation is faster than direct representation**

The second reason why the indirect way of representing evolved and not the direct way is that the indirect way is much faster. As argued before phenomenal experiences are immediate presentations. Even though qua representation they are indirect, qua presentation they are direct and it cannot get any faster than that. Deducing that someone has good reproductive features from the perception of — say — a symmetrical face takes much longer than seeing that one is beautiful.

**Evolution’s shorthand**

Natural selection does not care about direct representation, nor about getting to know what actually is represented by a given phenomenal experience. Natural selection only is interested in what works: a quick representation of a face as beautiful works better if it is linked to the appropriate reaction, than any direct representation of fertility, since it requires less background knowledge and is an immediate presentation. Those who did not respond in the appropriate manner evidently did not pass on their genes. Phenomenal consciousness then (at least in the case of experiences of beauty evoked by faces) works fast and contains sufficient information for the organism to survive and reproduce, even though the represented properties are not directly represented and leaves out any explicit information about the reproductive features of a conspecific. Phenomenal experiences, when used as representations, provide us (and other
animals) with a fast way to pick up highly complex information in a glimpse. This is why I call phenomenal consciousness **evolution’s shorthand**.

### 4.1 two other examples

One can imagine all other kinds of examples where phenomenal representation of the environment (including the organism itself) is preferred by evolution over another, slower but maybe more precise way of representing. Other empirical studies have to show whether these examples are indeed correct. I will discuss three unsupported but what I believe to be plausible examples to clarify the evolutionary shorthand theory of phenomenal consciousness.

**Predators**

In the environment in which humans evolved it would be a good thing to directly represent a predator as something that is threatening. A direct recognition and an **immediate response** are essential to this. This does not mean that there are no other ways of representing a tiger coming out of the woods. One can linguistically represent the situation and the response as follows: “that moving thing over there coming out of the woods, is an animal that is probably a feline, and if it is hungry it will try to eat me, which is bad for the survival for my genes and myself, hence I should run, hide, or kill it.” By the time the observer is halfway his reasoning, he will be the tiger’s dinner. Here too, the fact that phenomenal representations are evolution’s shorthand is evident: it is much faster and therefore evolutionary more advantageous than a complex linguistic representation.

**Gathering food**

The fact that phenomenal experiences are evolution’s shorthand of representing complex information can also be made plausible by considering how we gather our food. We do not wander around with our mouths open hoping an apple or a piece of meat gets in. In order to get our food we cannot just interact in a passive manner like some single-cell organisms can. In order to get our food we have to **actively explore** our environment. But our environment is a complex system and everywhere we go huge amounts of information are available. In chapter ten we saw that the color of an object is partly determined by the reflective properties of its surroundings. This might give us a hint of how food detection based on experiences of color works. Suppose we are confronted with an environment in which many different kinds of vegetation are present. It makes very good sense that we then can distinguish the ripe tomatoes from those that are not ripe yet, and that we can do this from dawn to dusk, i.e. that we can do this under different kinds
of lumination, which in laboratory conditions do change the color of an object placed out of a context. Since in natural circumstances the lumination changes for an entire scene, the phenomenon of color constancy plays an important role in distinguishing the ripe tomatoes from the unripe and the background. But we are not aware of the properties neither of the light nor of the reflectance properties of the surface of the tomatoes and their surroundings. We use our experience of color: the ripe tomato *pops out* and immediately draws our attention (that is, for color normals under every natural light condition).

There is some supported for this example. The popping out of red objects in a green environment is a recent phenomenon by evolutionary criteria. Most animals are dichromats, lacking long-wave sensitive cones. Their system consists of short- and middle-wave sensitive cones. This is commonly referred to as the *ancient system*. Some 30 million years ago there was a co-evolution of fruits in trees that became big enough for primates to be worth eating (and too big for birds) and the trichromatic color vision of those primates. The genes that are responsible for coding both the long- and middle-wave sensitive cones have 96% of their code sequence in common and are located close to each other on the X-chromosome. This makes it both very plausible that the long-wave sensitive cones were split into long- and middle-wave sensitive cones and that this split is a recent one (McIntyre 2002: 31-33).

**Proprioception**

The fact that the content of phenomenal experiences can represented in another, but more time-consuming manner is illustrated by persons who lack proprioception. There are few individuals that have no inner experience of how their bodies are positioned. To get this information, they have to look where their limbs are. Not until they have this information can they walk, and even then it is heartbreaking to see how severely impaired their movements are. The lack of proprioception shows that the same information can be represented in another way and that it can be used for the same purpose, but that the representation is no longer a direct presentation, and the use of this information requires a time-consuming processing of the information.

**4.2 affordances**

The account of phenomenal consciousness as evolution’s shorthand is somewhat similar to James Gibson’s (1966) account of *affordances*.

When the constant properties of constant objects are perceived (the shape, size, color, texture, composition, motion, animation, and position relative to other objects), the observer can go on to
detect their *affordances*. [...] I mean simply what things furnish, for good or ill. What they *afford* the observer, after all, depends on their properties. (Gibson 1966: 285)

As is clear from the previous chapter, I do not believe that objects really have colors, for these are secondary properties, so my account differs from Gibson’s in that I defend the position that we can use our own *experiences* as indicators of the primary properties of the objects in our environment. However, since we project our phenomenal experiences onto the objects out there (hence the qualification secondary property of an object) and this is done reliably so (a system that would not do that would not survive in evolution by natural selection) we still can classify this as detecting the affordance of the object. In the case of the ripe tomato we perceive its edibility *via our experience of red*. If we had to pickup this information by representing the edibility *as edibility*, then we should have knowledge about the function of eating, which is evolutionary impossible to have developed before the actual eating. Of course it is also possible to represent the edibility of a ripe tomato in another way, as protanopes and deuteranopes have to do; instead of the color, one can use the shape to detect where it is, and can then use the brightness of the color the tomato has and also the way it feels to evaluate its edibility. Needless to say that it takes more time to do so. The facts that it can be done and that it takes more time support my claim that phenomenal presentations are used as shorthand to represent information that can also be represented in another way. Instead of immediately seeing that the tomato is edible, one can also get this information by detecting where the tomato is and consequently squeezing it.

5. **some speculations about the experiences of beauty evoked by landscapes and by art**

The evolutionary shorthand theory can account for why presentations such as the phenomenal experiences of beauty evoked by faces evolved. What about experiences of beauty evoked by other *natural* and by *nonnatural* objects and events? Can the evolutionary shorthand theory also provide us with an explanation of these phenomenal experiences? First, according to the evolutionary shorthand theory phenomenal experiences are presentations that can be used as representations. Having said that, nothing is said about any possible restriction to what it can be used to represent. It is because of the fundamental possibility for any presentation to represent anything, that synesthetic representations are different representations and not misrepresentations. Even though I have argued that selection is not blind as to what is represented, this fundamental possibility of presentations to represent anything implies that presentations might just as well be used to represent objects that evidently have no evolutionary function, or those in which we do not immediately see what it is that is represented that is of evolutionary importance. These two might even be related.
Experiences of beauty are also used to represent features of natural objects other than those of conspecifics. For example experiences of beauty evoked by landscapes might represent that the environment is one that would be very beneficial for our ancestors: a beautiful landscape might indicate food, water and shelter against predators and other natural threats, as argued by Jay Appleton (1975: 68-74).

Experiences of beauty evoked by art might be related to this, in the way that we appreciate those properties in works of art that are abstracted from natural objects and scenes. Many works of art, for example, contain a high degree of symmetry, which we saw earlier was a feature of healthy bodies.

This is of course very speculative. There is a huge gap in the empirical data that would support claims like these. However, that is not relevant to the point I want to stress here: The important thing is that the evolutionary shorthand theory is in principle capable to explain why different natural and nonnatural objects can evoke similar phenomenal experiences and still be an evolutionary and representational account of phenomenal consciousness.

These conclusions drawn from the case study of experiences of beauty concludes my presentation of empirical data supporting the evolutionary shorthand theory of the phenomenal mind as explained in chapter two.

In the last chapter of this dissertation I will look at three objections to any representational theory of phenomenal consciousness. I will argue that the evolutionary shorthand theory of the phenomenal mind can counter these arguments. Besides defending the evolutionary shorthand theory of the phenomenal mind one further conclusion will be drawn based on the empirical data used in the third argument against representationalism: contrary to what one might expect phenomenal representations are not detailed representations.
CHAPTER TWELVE
THREE THREATS TO REPRESENTATIONALISM REFUTED

A representationalist theory of the phenomenal mind – even if it does not accept that all phenomenal states are representational – should address the following three problems. All arguments apply to those phenomenal presentations that are used as representations. First of all, it might be argued that a representational account of phenomenal consciousness loses consciousness, for it tries to reduce phenomenal experiences to representations. But what we know of representations in general – e.g. photographs, movies and natural representations like smoke representing fire – does not tell us anything about phenomenal consciousness. An account that reduces phenomenal experiences to representations loses consciousness.

Secondly, if the phenomenal experiences are used as representations and they supervene on brain states, do they represent the same as the neural states? If the answer is yes, this would mean that there is a problem of overrepresentation: anything that is represented by the phenomenal states is already represented by the brain state. This then, would amount to conscious inessentialism, as described in chapter one. This position is defended by Max Velmans (2000). This argument can be interpreted as a second reductionist threat: phenomenal representations can be reduced to neural representations.

Thirdly, there are theories that are based on the assumption that there are no internal representations at all (neither neural nor phenomenal). We can perform any action without any internal representation. The outside world functions as its own representation. This is what I call representation inessentialism. A variant of this theory is defended by Kevin O’Regan and Alva Noë (2001). They present empirical data to support their theory. Any representationalist theory of the phenomenal mind has to take into account these data.

In this chapter I will show that the first reductionist threat is no threat at all for phenomenal representations create intensional contexts. Furthermore, both inessentialisms are incorrect. In short, the evolutionary shorthand theory demonstrates that conscious inessentialism is wrong, for it falsely assumes that a phenomenal representation represents the same as its subvenient neural representation. Representation inessentialism is incorrect for the following reason: it heavily draws on refuting the representationalist account by showing that the alleged internal
representation is not a detailed copy of the outside world. Empirical data indeed support this claim. However, this does only refute a representationalist theory of the phenomenal mind if a theory would accept such a claim. At least the evolutionary shorthand does not.

1. **rebuttal of the first reductionist threat: losing consciousness**

A common issue concerning any representational account of phenomenal consciousness has a positive and a negative interpretation. The positive view is the following: if we can develop a representational account of phenomenal consciousness, then the problems of phenomenal consciousness can be solved, for we understand in general how one thing can represent another. All we have to do is to show that phenomenal experiences are representations. The negative view is the following: given the strategy proposed by the positive view, we are reducing phenomenal consciousness to representational states in general. But representational states in general are not phenomenally conscious. Hence, by reducing consciousness to representational states we eliminate consciousness. This then, would be an argument to accept conscious inessentialism (Cf. chapter one, section two).

In my view the positive view is too optimistic and the negative view to pessimistic. We need to show that we can give a representational account of phenomenal consciousness, without running the risk of losing it. I believe that the evolutionary shorthand theory meets this demand, for three reasons. First, and this may already be enough, phenomenal experiences are not always representations. Hence, phenomenal experiences are not always intentional states, which implies a general theory of intentionality will not be able to explain all the aspects of phenomenal consciousness.

Secondly, as I already argued, phenomenal experiences are shorthand, and this is necessarily so. For a nonconscious system to perform just as we do, it would first need to know many things about reproduction and evolution before it can react in an appropriate manner to nonphenomenal representations like “this is a predator and it’s coming toward you.” In order to both understand this and react as fast as we do, this implies that the system has to be complex at the start of its evolutionary development; it needs to have stored knowledge about predators and it has to process the information very fast. This is not how evolution via natural selection work, in evolution things go from simple to complex, not vice versa.
Thirdly, even if we look only at those situations in which phenomenal presentations are used as representations, we do not run the risk of losing phenomenal consciousness, for phenomenal representations create *intensional contexts*, which do not allow for substitution – hence do not allow for reduction – of phenomenal representations to nonphenomenal representations. For any intensional context goes that, even though we know from a *third person* perspective that two representational states represent the same object, property or event, we are not at liberty to substitute the representational states in their *first person* context, for this might change the truth value.

If we accept that phenomenal experiences are projected back on objects, generating secondary properties, and that this is an *indirect* way of representing primary properties, then it is legitimate to conclude that if phenomenal presentations are used as representations of properties of natural objects, all such phenomenal representations create intensional contexts, in a way similar to the creation of intensional contexts by experiences of beauty evoked by facial features.

This shows that the reductionist threat is no real threat at all, for there is no way we can substitute a linguistic representation for the phenomenal representation, for the person having a certain experience (that is representational) might not know what the primary properties are that are represented. Hence, we do not run the risk of reductionism.

Before I turn to the second reductionist threat, I first want to point out a difference between the way *intensionality* plays a role in my argument and the role it plays in an argument of Michael Tye. Tye presents an account of how perceptual sensations can represent. I agree with this account as far as perceptual sensations are a subclass of phenomenal experiences. Tye, however, wants so show that *all* phenomenal states are representational, hence he has to show that even afterimages are representational, a claim I deny. He uses the intensional character of phenomenal states to show that afterimages represent that something has a certain color $F$ and a certain shape $G$. And even though $F$ might be red, which is the color of most fire engines, something is represented as red and not as having the color of most fire engines: experiencing something as red creates an intensional context (1995: 107). I agree that it does. However claiming that this applies to all phenomenal states is begging the question against those that reject representationalist theories of the phenomenal mind and those that argue that not every phenomenal experience is a representation. It is only after we have established that a certain phenomenal presentation is used as a representation, that we can conclude that it creates an intensional context. In the case of
afterimages it is clear that there is not something that is represented as being colored or having a shape. If one has an afterimage, one simply experiences a color and a shape and it is very clear that there is no object that has this color and shape. Hence, there is nothing that can be substituted for what is represented, for there is nothing that is represented.

2. rebuttal of the second reductionist threat: conscious inessentialism via overrepresentation

There is a second reductionist threat. This threat comes from the conceivability of the truth of overrepresentation: the phenomenal representation can be reduced (and thereby eliminated) to the neural representation on which it supervenes, for both represent the same.

If phenomenal presentations are used as representations, why then are they not superfluous? If I am right, that phenomenal presentations supervene on neural states that are also representational, doesn’t this lead to what I call the problem of overrepresentation? The argument would be the following. Both the neural and the phenomenal state seem to represent the same features of the same referent. If this indeed is the case, then all the representational work is already done by the neural representation. What, then, is the additional function of phenomenal experience? Evidently, the problem of overrepresentation is closely related to the problem of causal overdetermination: if our brains do all the causal work what, then, does phenomenal consciousness do?

Both the problem of overrepresentation and causal overdetermination amount to either one of the following positions. First, one might defend epiphenomenalism with respect to phenomenal consciousness. Secondly, there is the related position in which one accepts the problems of overdetermination and overrepresentation as a paradox, like Chalmers’ paradox of phenomenal judgment (1996b; Cf. chapter three) or Velmans’ causal paradox (2000), without embracing epiphenomenalism. I believe that in both cases they are merely denying that their views lead to epiphenomenalism, for their theories clearly implicate that – if they are right – phenomenal states are neither causally efficacious, nor do they represent anything over and above the physical representations with which they are correlated.

Thirdly, one might conclude that phenomenal states can be reduced to neural states and that eliminativism is correct. If the phenomenal representations encodes the same information about a referent as the neural representations on which they supervene, and do not do anything that is
not already done by these states, we can reduce the phenomenal states to the neural states, hence we are losing consciousness.

Either of these positions can be viewed as forms of conscious inessentialism. I will argue that the information contained in the phenomenal state represents different features from the referent than the neural representation does. This shows that there is no overrepresentation, hence, overrepresentation is no reason to accept conscious inessentialism. The further implication is that we do not have to choose between the three options I just mentioned. The case study of experiences of beauty evoked by faces of conspecifics can be used to demonstrate that there is no problem of overrepresentation.

What do the neural representations evoked by a face represent? First, they represent the different frequencies of the light that was reflected from the face, even though we are not consciously aware what those frequencies are. I believe that this already is enough to show that the neural and the phenomenal representation differ with respect to their contents. Secondly, the brain reconstructs the features of the face, based on the features of the light that stimulates the retina. What is represented neurally are the shape and size of the face, not whether it is beautiful or not, even though these are of course necessary for the experiences of beauty. It is the integration at the phenomenal level that results in experiences of beauty.

What is represented at the neural level are some of the primary qualities of the thing that is observed, like the shape, the size, the kind of light that it reflects, while the phenomenal representations represent those features that are of interest to the representing organism: the health or fertility of a possible mate or whether something is edible or not. Clearly, those properties are not represented at all at the neural level.

Another example is that the neural representation of a ripe tomato is a representation of the reflectance properties of the tomato (and its surroundings), while the phenomenal representation does not but it does represent that it is edible, via the phenomenal experience of red.

Yet another example might be the phenomenal experience of fear, while perceiving a tiger. Again, the neural representation is just one of the shape, size, movement etcetera of the object, and of the properties of the light that is reflected by it. The fact that it is a predator is not represented at the neural level. The neural representations on which phenomenal representations supervene but
represent different aspects of the represented object. This means that phenomenal representations cannot be reduced to their subvenient neural representations. This, in turn, implies that the problem of overrepresentation does not arise, for there simply is no overrepresentation. Hence, we are not forced to accept conscious inessentialism and choose between the epiphenomenal and eliminativist version of this position.

3. rebuttal of representation inessentialism

I have just argued that the two reductionist threats to a representational theory of the phenomenal mind are in fact no threats at all. The two arguments started by accepting the claim that phenomenal experiences are representations, and went from there to show that reductionism might follow.

Recently, a more fundamental criticism of any representationalist account of phenomenal consciousness has been put forward by Kevin O’Regan and Alva Noë (2001): we do not need representations at all to explain that we are (visually) conscious of something. We could call this representation inessentialism: representations are inessential to our conscious actions (at least internal or mental representations are). O’Regan and Noë present empirical data that support their view.

I will now address this question and show that from an evolutionary point of view this account only applies to behavior (and not action) triggered by single cues. Since it is doubtful that this behavior is guided by consciousness, it is also doubtful that the account given by O’Regan and Noë is an account of consciousness at all. Furthermore, the evolutionary shorthand theory can handle the empirical data that O’Regan and Noë present.

O’Regan and Noë claim that a representational account of consciousness cannot be correct:

Many current neurophysiological, psychophysical and psychological approaches to vision rest on the idea that when we see, the brain produces an internal representation of the world. The activation of this internal representation is assumed to give rise to the experience of seeing. The problem with this kind of approach is that it leaves unexplained how the existence of such a detailed internal representation might produce visual consciousness. [...] We propose that seeing is a way of acting. It is a particular way of exploring the environment. Activity in internal representations does not generate the experience of seeing. The outside world serves as its own, external, representation. (2001: 939)

O’Regan and Noë focus on visual consciousness. According to them “the experience of seeing occurs when the organism masters what we call the governing laws of sensorimotor
contingency.” (2001: 939) To see how this is an account of consciousness, we first have to answer two questions: What are sensorimotor contingencies? And What are the laws that govern them?

Consider a car that passes by. Most people will have different kinds of experiences: you will see the car passing by, you will also hear it, and if the car passes by closely you will also smell it and feel the wind it produces. Some of the light reflected by the car will fall on your retina, and the pattern on your retina will change as you or the car moves. These changes are governed by certain laws. Something similar goes for the auditory experience of the car passing by: here it is not light that is the stimulus, but oscillations in the air. These obey different laws than light does. Also, since it is not the eyes that are affected by the stimulus, different laws apply when the ears move relatively to the car, then in the case of the eyes that move relatively to the car. The difference between seeing the car and hearing the car then are governed by laws that are determined by the exploring sensory modalities and what they are exploring (2001: 941). These changes are what O’Regan and Noë call sensorimotor contingencies. Sensorimotor contingencies come in two kinds: those that are determined by the sensory modality (corresponding to the traditional notion of “sensation”) and those fixed by the character of the object that is being tracked (corresponding to the traditional notion of “perception”) (2001: 943).

Each sensory modality, then, has its own sensorimotor contingencies and the laws that come with them. It is these laws that play an essential role in the theory of consciousness of O’Regan and Noë. They focus on vision. Two basic conditions have to be satisfied for vision: An animal has to explore the environment in a way that is governed by both kinds of sensorimotor contingencies, and it has to do so by actively exercising its mastery of the appropriate laws (2001: 943). This means that the animal must have knowledge of these sensorimotor contingencies and of the appropriate laws:

Visual perception can now be understood as the activity of exploring the environment in ways mediated by knowledge of the relevant sensorimotor contingencies. And to be a visual perceiver is, thus, to be capable of exercising mastery of vision-related rules of sensorimotor contingency. (2001: 943)

It seems that the mastery of the rules of the sensorimotor contingencies is enough, and that we do not need an additional representation of the world to guide our actions. O’Regan and Noë support their view by presenting empirical data about change and inattentional blindness. They use these data to argue against the intuition that we have a detailed internal representation of everything we look at.
They support their view that this is not the case by the empirical data coming from change and inattentional blindness experiments. In change blindness subjects are looking at a scene in which gradually a radical change is made. For instance the shirt of the person in the scene changes color, or a road slowly appears in a corn-field. This changes is not noticed. Even more shocking to most people is inattentional blindness. Subjects are shown a movie in which two teams are passing basketballs at each other. One is instructed to count the times the white team passes the ball. When performing this task approximately half of the observers fail to notice the woman in a gorilla suit that walks into the middle of the display, thumped its chest, and leaves again (Simons & Chabris 1999). From these experiments it follows that “Our intuition that we richly represent the visual details of our environment is illusory.” (Simons 1997: 501) This is exactly what O’Regan and Noë argue for.

The results of the experiments showed that in many cases observers have great difficulty seeing changes, even though the changes are very large, and occur in full view – they are perfectly visible to someone who knows what they are. Such results are surprising if one espouses the view that we should “see” everything that we are looking at: It is very troubling to be shown a picture where a change is occurring repetitively and in full view, without being able to see the change. The experience is quite contradictory with one’s subjective impression of richness, of “seeing everything” in the visual field. However, the results are completely coherent with the view of seeing which is being defended here. (2001: 954)

The question is whether these data are also enough to get representationalism into serious trouble. This is what O’Regan and Noë seem to think.

It is generally thought that somewhere in the brain an internal representation of the outside world must be set up which, when it is activated, gives us the experience that we all share of the rich, three-dimensional, colorful world. (2001: 939)

It will be clear that I do not agree with O’Regan and Noë, for I defend a representationalist account of phenomenal consciousness. I will present three points of criticism. The first concerns a problem in the theory itself, the second is a defense of the need of representations for complex situations as is argued for by Kim Sterelny (1999). The third problem seems to be the most serious: empirical data refute any representationalist account. The last part of this chapter will be dedicated to refuting the argument that O’Regan and Noë get from these data.

**Point of criticism # 1**

The first point of criticism is one that in my opinion is both obvious and fatal: their theory lacks the same explanatory power they claim the representational theories are supposed to lack. O’Regan and Noë accuse theories that accept internal representations of having only found the
neural correlations of consciousness at best, and since correlation is not causation these findings do not explain why activation of these representations should give rise to consciousness. The question why these correlates – say 40 Hz synchronous oscillations – produce consciousness remains unanswered, and “the problem of consciousness would simply be pushed back into a deeper hiding place[,]” (2001: 940) This is misleading. The firm corroboration of 40 Hz hypothesis (Cf. chapter nine) does not push consciousness deeper into a hiding place, but puts a spotlight on it: this is where we have to search for an explanation. It rules out other explanations, hence we do not have to waste time any more to – say – proposed explanations of consciousness at the quantum level. So, even though correlation is not explanation, searching for the neural correlates of consciousness is not as futile as O’Regan and Noë want to us to believe. The real problem with their own theory is that it does not explain why consciousness arises in some creatures, for O’Regan and Noë themselves claim that their account of consciousness applies also to nonconscious systems such as a missile guidance system (2001: 943). But if an account applies both to conscious and nonconscious systems, then evidently the account cannot explain consciousness.

**Point of criticism # 2**

The second point of criticism is derived from Kim Sterelny (1999). Sterelny argues against the so-called theory of situated agency, which claims that intelligent action is possible without internal representations. Part of this theory is that the information is stored outside the organism (1999: 204). Even though O’Regan and Noë primarily give an account of visual consciousness and not of intelligent action, the argument Sterelny makes against the situated agency theory also holds against the sensorimotor theory.

The argument is straightforward and is implied by the evolutionary shorthand theory of phenomenal consciousness. Many organisms are agents, and to be an agent is more than just reacting to the environment based on single cues. Some animal behavior can be described this way. For instance ants and bees do not see that a nestmate is dead. The way they recognize that they are dead is by the detection of the oleic acid that is produced by the decay of the dead nestmate. After detection they clean their nest, hence the disposal of dead nestmates is driven by a single cue. Much insect behavior can be described this way. Behavior based on single cues is prone to parasitism. If an organism produced the same chemical cue as young insects do, they can enter the nest and will be fed by the adult insects (1999: 206). However, if the environment becomes complex and many different objects are the causes of the same stimulus (the same cue),
then one needs to choose between where one reacts to, and this means one needs to develop preferences and desires. But this is not possible without having representations of one’s environment, for choosing implies planning: not only does one need to perceive how the world is now, one also needs to represent the world how it will be if X rather than Y is pursued. Having preferences and planning one’s actions, requires representations of both the desires and ways to fulfill them (1999: 211, 214).

Representing the world as it will be, evidently cannot be done without representations, hence knowledge of sensorimotor contingencies will not be enough. O’Regan and Noë’s theory might be enough to account for the behavior of insects, that seems to be single cue driven (at least some of their behavior is). Where such insect behavior is concerned, the demands that (a) than changes in cues are followed by a behavior response, and (b) that the mastery of the sensorimotor contingencies shows in the behavior of the insect (2001: 943) are met. However, it is precisely in these cases that the other mind problem becomes a real problem (as opposed to the question of whether other humans do have minds): do insects have consciousness? A positive answer to this question is by no means evident, for all we have to go on in determining whether other animals are conscious is the similarity of their behavior and brain, compared to ours. Sverre Sjölander (2001) uses these criteria to evaluate animal consciousness. He argues that other mammals and birds probably also have consciousness, for their brains are similar and so are their actions. A cat for instance will chase a mouse when it disappears from view; the cat seems to have a representation of what is behind its field of view. Reptiles, like snakes do have brains that are dissimilar to ours. So is their behavior. If a mouse flees behind a rock, a snake stops looking at it, does not look for it, and does not behave as if it knows that there is a mouse behind the rock. Sjölander, then, argues that we cannot be sure whether reptiles do have consciousness. Since insects differ even more from us in brain and behavior demonstrating that they do have consciousness is an even harder problem.

Since it is hard to determine whether insects have conscious experiences, and O’Regan and Noë’s theory seems to be adequate only to explain the information pickup and behavior of animals like insects, it is hard to see how theirs can be a theory of consciousness. A theory of consciousness should focus on those animals we are sure of that they are conscious beings. This means that we should focus on humans (and other mammals and birds). In our own case the environment is of such complexity that one needs make choices and plan ahead, which makes representations of what one wants and how the environment will chance according to one’s actions.
**Point of criticism # 3**

My third point of criticism concerns the claim that change and inattentional blindness show that representationalism cannot be correct, since representationalism adheres to the view that we have an inner and detailed representation of the outside world. Clearly, we do not have such inner detailed representations, for then we would notice such a substantial change in our visual field as a gorilla walking into a basketball field. However, having an internal representation does not imply that this representation is as detailed as the outside world. What would the evolutionary function be of representing the outside world in every detail? I already argued that evolution does not care about what is a direct representation and the same goes for a detailed representation. Apparently we can go about in the world perfectly well using representations that are not as detailed as the world around us is. From an evolutionary viewpoint it would even be implausible if we had such representations, for then we would have too much information about the objects and events in our environment: we would have to evaluate every detail, which probably would be a time-consuming process. It makes much more sense to consciously represent only those features that are of interest to the organism, with the option that a shift of attention can result in seeing details where one did not before. The data provided by change and inattentional blindness experiments show exactly that evolution indeed did not work that way.

The three objections against representationalist theories of the phenomenal mind put forward in this chapter have all been met by the evolutionary theory of the phenomenal mind. While theories of Tye and Dretske might not have an answer to the first objection, the view that phenomenal experiences are not always used as representations neutralizes that objection. Furthermore, where Velmans gets himself into trouble, viz. by his claim that it is plausible that the neural and the corresponding phenomenal representation represent the same, I have shown that these representations have different content. The objection of O’Regan and Noë, at last, does not stand up whatever representationalist theory of the phenomenal mind one defends.
In this dissertation I have argued against those who believe for different reasons that scientific investigation will not help to solve the problem of phenomenal consciousness. The chapters in which I present empirical data show otherwise. These chapters form the most important argument in favor of the view that science can help. The evolutionary shorthand theory provides a view on the phenomenal that is not new: in the long tradition of thinking about mind and body all of the different aspects have been proposed. In this dissertation, though, all the different aspects are the result, not of armchair philosophy, but of the combination of empirical data from different scientific disciplines. What follows is the characterization of the phenomenal mind, as presented in chapter two:

Phenomenal experiences form a heterogeneous set of causally efficacious internal and direct presentations. In most normal conditions they are used to represent real properties of external objects or bodily states in a nonconceptual, indirect though reliable way, approximately projecting them back on these states with which they causally covary. In other words: they represent or track those properties. The normal conditions are those conditions in which the experiencing organisms evolved. Phenomenal representations are a species-specific evolutionary shorthand for complex information about the environment or the body that – in principle – could also be represented in another less economic way. It is precisely this economic manner of representing, which makes phenomenal consciousness a feature that is selected for in evolution by natural selection. In all other conditions phenomenal experiences are either just presentations not representing anything (e.g. afterimages), or they are used to represent features that are not found in the environment in which the experiencing organisms evolved (e.g. the beauty of art). Since phenomenal experiences were selected for their representational abilities, experiencing organisms have the natural inclination to interpret any phenomenal experience as representational, even if it is merely a presentation.

The most important claim is that usually phenomenal experiences are presentations that are used as representations. This makes the evolutionary shorthand theory a special representationalist theory, for it does not allow for a reduction of the phenomenal experience when it is used as a representation to nonphenomenal representations.

Furthermore, having demonstrated the presentational character of phenomenal experiences, the evolutionary shorthand theory of the phenomenal mind can abstain from strange claims that
afterimages (and the like) are misrepresentations of – say – the wall one is looking at. This distinguishes the theory from other representationalist theories.

Another important aspect of the above characterization is that the issue between internalism and externalism about phenomenal content is scientifically settled in favor of internalism, which is consistent with the common sense belief that it is the brain that determines the phenomenal experience and not something out there.

_The problem of phenomenal consciousness revisited_

As I mentioned in chapter one, the heterogeneity of phenomenal consciousness makes it unlikely that a solution to the problem of phenomenal consciousness can be presented in a single article or book. Hence, I do not claim that the aspects mentioned in the characterization above summarize all there is to phenomenal consciousness. There are many questions about phenomenal consciousness that remain. The evolutionary shorthand theory, however, does entail some suggestions for the remaining answers.

The main question is that makes up half of the problem of phenomenal consciousness: _why do we have phenomenal consciousness given our physical make-up?_ (Cf. chapter one) The evolutionary shorthand theory does not provide an answer, for the relevant data are just not known yet. However, the discussion of dreams and how to determine whether one is dreaming showed that the 40 Hz hypothesis of Crick and Koch no longer is a hypothesis and points into the direction in which the study of consciousness has to look.

Also, with respect to the other half of the problem of phenomenal consciousness, viz. _why did phenomenal consciousness evolve?_ Not all questions are answered. An important, unanswered question here is the question how phenomenal consciousness came into being, for it is evidently not the case that a conscious creature came into being with the heterogeneous set of conscious experiences we have today. Here we can speculate starting from what has been said in chapter twelve. In arguing against the sensorimotor account of consciousness, I claimed that this theory might solely apply to single-cue based behavior like that of an insect removing a dead conspecific from the nest only if it produces a certain chemical. In cases like these it is not clear whether it is just a chemical reaction with no phenomenology involved or whether the dead insect is removed from the nest because it smells bad. The latter scenario is at least _conceivable_. But as I argued in chapter three, conceivability does not lead us to knowledge. For now, I take these questions as
problems that need to be solved in accordance with the evolutionary shorthand theory of the phenomenal mind.
INLEIDING

Dit proefschrift gaat over de wetenschappelijke studie naar bewustzijn. Als we kijken naar de bijdragen aan recente conferenties die als thema de wetenschappelijke bestudering van bewustzijn hebben, dan kan men niet ontsnappen aan de indruk dat er consensus lijkt te bestaan over de mogelijkheid dat de wetenschap uiteindelijk een oplossing kan geven voor het probleem van fenomenaal bewustzijn zoals dat in de cognitiefilosofie bestaat.

Ondanks deze groeiende consensus en de opeenstapeling van nieuwe gegevens over het bewustzijn, is er toch een aantal filosofen die om verschillende redenen meent dat een wetenschappelijke bestudering van bewustzijn niet mogelijk is. In dit proefschrift argumenteer ik tegen een aantal van deze posities. Het voornaamste argument tegen diegenen die menen dat een wetenschap van bewustzijn onmogelijk is, is om te laten zien dat het wel degelijk kan. Daartoe presenteer ik de *natural method* van Owen Flanagan. Door deze toe te passen op een aantal soorten bewuste ervaringen, volgt wat ik de *evolutionair steno theorie van fenomenaal bewustzijn* noem. Deze theorie geeft een antwoord op de vraag *waarom fenomenaal bewustzijn geselecteerd werd in de evolutie*, het is een snelle manier om informatie op te pikken die van belang is voor het voortbestaan en de reproductie van het bewuste organisme. Dit zal niet als een verrassing komen omdat dit immers geen nieuw idee is. Mijn doel in deze dissertatie is dan ook niet om een nieuwe theorie te presenteren, maar een die ondersteund wordt door recente empirische data.

HOOFDSTUK I

TWEE AANNAMEN EN TWEE VRAGEN

Ik houd erg van *Chimay Grande Réserve*. Soms maakt dat dat ik er teveel van drink wat als gevolg heeft dat ik de volgende dag een vreselijke kater heb. Volgens mij is dit het kortste argument wat gegeven kan worden om te laten zien dat fenomenale ervaringen bestaan en dat ze causaal interacteren met de rest van de fysische wereld. Als we verder de causale geslotenheid van de fysische wereld accepteren en causale overdeterminatie uitsluiten, dan volgt dat fenomenaal bewustzijn ook een (bepaald soort) fysische eigenschap is.

In hoofdstuk één doe ik twee dingen. Ten eerste zet ik de twee vooronderstellingen uiten die ik accepteer in het hele proefschrift: *qualia realisme* en *breed fysicalisme*. Ten tweede geef ik de twee vragen die samen het probleem van fenomenaal bewustzijn vormen.
AANNAME ÉÉN: QUALIA REALISME
Over qualia realisme kan ik vrij kort zijn: het is iets om een mens te zijn, in tegenstelling tot een robot of een zombie. Een zombie wordt in de cognitiefilosofie gedefinieerd als een fysisch en functioneel identiek duplcataat (in een parallel fysisch en functioneel identieke wereld) die echter geen fenomenale ervaringen heeft. Mijn zombie-broer gaat in de zombie-wereld wel naar het café als ik dat hier doe, zal ook langzaamaan moeite met praten, denken en lopen hebben, maar zal daarnaast geen prettige ervaringen hebben. Gelukkig voor mijn zombie-broer heeft hij de volgende dag ook geen hoofdpijn. Het is dus iets voor een mens om bier te drinken, terwijl dat voor een zombie of robot niet zo is. Dit aspect is wat fenomenaal bewustzijn is, en daar gaat het om in dit proefschrift.

AANNAME TWEE: BREED FYSISALISME
Als je zegt dat iets echt is, dat iets bestaat, dan moet je zeggen wat je daarmee bedoelt. In mijn visie bestaat iets enkel dan als het tot het brede fysische domein behoort. Hiermee wordt de vraag direct: wat is het brede fysische domein? Dat is dat domein van eigenschappen die ofwel zelf fysisch zijn of door fysische entiteiten of eigenschappen geconstitueerd worden. Biologische eigenschappen zijn zelf geen fysische eigenschappen, maar worden wel door fysische entiteiten – moleculen – geconstitueerd of gerealiseerd. Een biologische eigenschap is daarmee een brede fysische eigenschap.

Breed fysicalisme is een vorm van monisme, omdat het stelt dat uiteindelijk alles wat bestaat geconstitueerd wordt door een onderste fysische laag: er is slechts één soort “spul” waaruit alles is opgebouwd. Je kan de vraag stellen of je dat nog wel fysisch moet noemen, maar dat is een kwestie van terminologie. Ik kies ervoor om dat fysisch te noemen.

Breed fysicalisme is ook een vorm van pluralisme, omdat het stelt dat hogere-orde eigenschappen (die geconstitueerd worden door lagere-orde eigenschappen) nieuwe eigenschappen zijn. De claim dat dit nieuwe eigenschappen zijn is een andere manier om te zeggen dat de hogere-orde eigenschappen niet gereduceerd kunnen worden tot lagere-orde eigenschappen. Breed fysicalisme is dus een non-reductionistisch fysicalisme. Dit wil echter niet zeggen dat er geen reductionistische verklaring mogelijk is van hogere-orde eigenschappen: precies omdat lagere-orde eigenschappen de hogere-orde eigenschappen realiseren, zijn deze nodig in een verklaring die aangeeft (1) wat de hogere-orde eigenschap doet, en (2) hoe dat gedaan wordt.
HET PROBLEEM VAN FENOMENALAAL BEWUSTZIJN

Mijns inziens is het probleem van fenomenaal bewustzijn niet *the hard problem*, zoals David Chalmers (1996a; 1996b) meent, maar gaat het om de combinatie van de volgende twee vragen:

(1) *waarom hebben we fenomenaal bewustzijn gegeven onze fysische structuur?* En (2) *waarom is fenomenaal bewustzijn gevormd?*

Men kan de eerste vraag interpreteren op de manier waarop Chalmers dat doet. De assumptie die men daarbij maakt is dat fenomenaal bewustzijn niet gerealiseerd wordt door fysische of breed fysische entiteiten of eigenschappen, maar dat het iets radicaal anders is. Als je de vraag op die manier interpreteert, dan moet je inderdaad concluderen dat de hedendaagse wetenschap het antwoord principieel schuldig moet blijven. In hoofdstuk vier toon ik aan dat Chalmers’ veronderstelling tot een *reductio ad absurdum* leidt, en dat deze interpretatie dus onjuist is.

Hoe moeten we de eerste vraag dan wel interpreteren? We moeten accepteren dat fenomenaal bewustzijn een natuurlijke, hogere-orde, biologische eigenschap is, en we willen weten hoe die gerealiseerd wordt. Op deze vraag geef ik slechts een gedeeltelijk antwoord. De reden daarvoor is dat het antwoord mede moet komen van de neurowetenschap, en dat die eenvoudigweg niet ver genoeg ontwikkeld is om op alle relevante vragen een antwoord te kunnen geven.

Op de tweede vraag geef ik wel een antwoord, waarvan ik hoop dat het een bevredigend antwoord is voor diegenen die realisten omtrent fenomenaal bewustzijn zijn en menen dat het causaal effectief is.

HOOFDSTUK TWEE

DE EVOLUTIONAIR STENO THEORIE VAN FENOMENALAAL BEWUSTZIJN

De inzet van het proefschrift is om het volgende aan te tonen.

_Fenomenale ervaringen vormen een heterogene verzameling van causaal effectieve interne en directe presentaties. In de meeste normale gevallen worden ze gebruikt om echte eigenschappen van externe objecten of lichamelijke toestanden op een niet-conceptuele, indirecte maar betrouwbare manier te representeren, waarbij deze representaties teruggereactiveerd worden naar een locatie nabij de toestanden waarmee ze causaal covariëren. Met andere woorden: ze representeren ofwel sporen met deze eigenschappen. De normale omstandigheden zijn die waarin de ervarend organismen_
evolueerden. Fenomenale representaties zijn een soort-specifiek evolutionair steno voor complexe informatie over de omgeving of het lichaam die – in principe – ook op een andere, minder economische, manier geregistreerd zou kunnen worden. Het is precies deze economische manier van reprezenteren die van fenomenaal bewustzijn een eigenschap maakt die geselecteerd werd in evolutie door natuurlijke selectie. In alle andere omstandigheden zijn fenomenale ervaringen slechts presentaties die helemaal niets representeren (vb. nabeelden) of worden ze gebruikt om zaken te representeren die we niet terugvinden in de omgeving waarin het ervarende organisme evolueerde (vb. de schoonheid van kunst). Omdat fenomenale ervaringen geselecteerd zijn vanwege de mogelijkheid om te representeren, hebben de ervarende organismen de natuurlijke neiging om elke fenomenale ervaring te interpreteren als een representatie, ook als het slechts om een presentatie gaat.

In de hoofdstukken zeven tot en met elf zullen empirische data aangedragen worden, waaruit deze beschrijving van fenomenaal bewustzijn volgt. Voordat ik die data presenteer zal ik eerst een aantal posities weerleggen die stellen dat een wetenschappelijke oplossing van het probleem van fenomenaal bewustzijn (nu) onmogelijk is.

Ten eerste is er de bedreiging van een versie van dualisme: indien fenomenaal bewustzijn niet fysisch of breed fysisch is, dan kan de hedendaagse wetenschap geen bijdrage leveren aan de oplossing van het probleem van fenomenaal bewustzijn, omdat de wetenschap zich beperkt tot het domein van het breed fysische. In hoofdstuk drie zal ik laten zien dat het recentelijk meest gebruikte argument voor dualisme - het voorstelbaarheidsargument - niet deugt, waarna ik in hoofdstuk vier zal laten zien dat de meest recent dualistische positie (namelijk die van Chalmers) zelf ook onhoudbaar is.

Met de dualistische bedreiging uit de weg zijn we er echter nog niet. Ten eerste is er de agnostische positie waarbij men ervan uit gaat dat we nu niet inzien hoe we het probleem van fenomenaal bewustzijn op moeten lossen en we dus agnostisch moeten zijn met betrekking tot het accepteren van monisme of dualisme. In hoofdstuk vijf zal ik laten zien dat het argument voor de agnostische positie voortkomt uit een te sterke eis aan een theorie over fenomenaal bewustzijn.

Ten tweede is er ook een bedreigend argument van de naturalistische, monistische zijde: we zijn cognitief te beperkt om de oplossing van het probleem van fenomenaal bewustzijn te begrijpen,

178
hoewel die oplossing in principe monistisch van aard is. We zitten in een positie die vergelijkbaar is met die van chimpansee's die kwantummechanica proberen te begrijpen. Het argument dat gebruikt wordt om deze visie te ondersteunen is het volgende. We hebben slechts twee methoden tot onze beschikking om het probleem op te lossen: introspectie en hersenonderzoek. Introspectie zal nooit laten zien dat we voor bewustzijn een lichaam nodig hebben, en als je hersens onderzoekt zal je nooit bewustzijn zien. We hebben geen idee hoe we het probleem op kunnen lossen. In hoofdstuk zes presenteer ik Flanagans argument tegen deze positie. Er is wel degelijk nog een methode om bewustzijn te bestuderen en tot antwoorden op bewustzijnsvragen te komen: de natuurlijke methode die inhoudt dat we andere methoden combineren, en dan met name fenomenologie, psychologie en neurowetenschap. Een voorbeeld wordt besproken om aan te tonen dat Flanagans claim correct is. In de hoofdstukken zeven tot en met elf pas ik deze methode toe om zo uiteindelijk tot de bovenstaande karakterisering van fenomenaal bewustzijn te komen. In hoofdstuk twaalf bespreek ik dan nog drie mogelijke objecties tegen een representationele theorie van fenomenaal bewustzijn, waarop de evolutionair steno theorie van fenomenaal bewustzijn een antwoord heeft.

HOOFDSTUK DRIE

VOORSTELBAARHEID: DE WEG DIE NERGENS HEEN LEIDT

Als we ervan uitgaan dat de wetenschap monisme accepteert, dan komt de belangrijkste bedreiging voor een wetenschappelijke oplossing van het probleem van fenomenaal bewustzijn van die vormen van dualisme die claimen dat fenomenaal bewustzijn naast niet te reduceren ook niet reductionistisch te verklaren is. Het gangbare argument is een voorstelbaarheidsargument: het is voorstelbaar dat er in een fysisch een functioneel identiek universum duplicaten van ons rondlopen die wel fysisch en functioneel exact hetzelfde zijn, maar die radicaal verschillen waar het fenomenaal bewustzijn betreft: ze hebben het of helemaal niet (zombie) of het is omgekeerd (omgekeerde tweeling). Ik hoofdstuk drie laat ik zien dat uit de voorstelbaarheid van dit soort scenario's geen conclusies met betrekking tot de onjuistheid van (breed) fysicalisme kan leiden.

Traditioneel wordt een aantal soorten mogelijkheid onderscheiden. De belangrijkste zijn logische, epistemische, ontologische en metafysische mogelijkheid. De verschillen zitten in de criteria waarmee beoordeeld wordt of iets als dan niet mogelijk is.

Logische mogelijkheid valt uiteen in formele en conceptuele mogelijkheid. Een uitspraak is formeel mogelijk als deze geen syntactische contradictie bevat (A & ~A). Een uitspraak is conceptueel
mogelijk als deze geen conceptuele contradictie bevat (een vrijgezel is getrouwd). Een uitspraak kan dus formeel wel mogelijk maar conceptueel onmogelijk zijn.

Epistemische mogelijkheid valt ook in twee versie uiteen. Iets is *slechts epistemisch mogelijk* als we geen informatie hebben die een bepaald scenario uitsluit. Iets is *positief epistemisch mogelijk* als we enige reden hebben om het scenario te geloven.

Ook met betrekking tot ontologische mogelijkheid kunnen we een onderscheid maken. Ten eerste is er *fysische mogelijkheid*. Iets is fysisch mogelijk indien er geen fysische wetten zijn die het uitsluiten. Men kan van mening zijn dat er naast de fysische weten ook nog andere weten zijn. Dit is de visie van Chalmers: de fysische en niet-fysische wetten samen vormen wat hij de natuurlijke wetten noemt. Daarmee wordt iets *natuurlijk mogelijk* als het de natuurlijke wetten niet schendt. Fysische wetten zeggen niets over niet-fysische entiteiten, waardoor het fysisch mogelijk is dat er engelen of ectoplasmatische entiteiten zijn. Dit is natuurlijk onmogelijk wanneer niet-fysische wetten het uitsluiten.

Dualisten moeten aantonen dat zombies *fysisch mogelijk* zijn. De conclusie die uiteindelijk getrokken wordt is dat fysicalisme onjuist is. Het gaat daarmee om een ontologische claim en om een claim die betrekking heeft op het domein van het breed fysische, wat dus betrekking heeft op ontologische mogelijkheid en dan op de variant daarvan die stelt dat er enkel fysische wetten zijn. In hoofdstuk drie laat ik zien dat geen van de andere varianten van mogelijkheid een route oplevert naar fysische mogelijkheid. Zo leidt logische mogelijkheid niet naar fysische mogelijkheid. Het is bijvoorbeeld logisch mogelijk dat zwarte gaten toegangs wegen zijn tot andere dimensies, omdat we geen formele noch een conceptuele contradictie kunnen ontdekken in dit scenario. We weten echter niet of dat er fysische wetten zijn die dat verbieden: we moeten meer weten om de stap van logische naar fysische mogelijkheid te maken.

Een argument tegen fysicalisme dat begint met de voorstelbaarheid van zombies, moet dan ook uitkomen bij de fysische mogelijkheid van zombies. Als daarin een tussenstap gemaakt wordt via een ander vorm van mogelijkheid, moet beargumenteerd worden hoe men van die vorm van mogelijkheid naar fysische mogelijkheid komt. Het argument tegen de anti-fysicalistische voorstelbaarheidsargumenten is precies dat dat achterwege blijft.

Het argument van Chalmers tegen fysicalisme gaat in het kort als volgt:
1. In onze wereld zijn er bewuste ervaringen.
2. Er is een logisch mogelijke wereld fysisch identiek aan de onze, waarin de positieve feiten omtrent bewustzijn niet gelden.
3. Daarom zijn feiten met betrekking tot bewustzijn verdere feiten van onze wereld, bovenop en naast de fysische feiten.
4. Daarom is materialisme onjuist. (1996b: 123)

Het is duidelijk dat Chalmers hier uit de logische mogelijkheid van een zombie wereld (stelling twee) een ontologische conclusie trekt (stelling drie) zonder daarbij aan te geven hoe die conclusie volgt. Chalmers’ argument is dus ongeldig.

HOOFDSTUK VIER
DAVID CHALMERS’ EIGENSCAPSDUALISME

Uit de ongeldigheid van een argument volgt niet dat datgene waarvoor geargumenteerd wordt ook niet klopt. In dit geval is het argument voor dualisme ongeldig, waarmee echter niet gezegd is dat ook de conclusie ongeldig is. In hoofdstuk vier laat ik zien dat ook de conclusie onjuist is het eigenschapsdualisme van Chalmers is onhoudbaar omdat het een contradictie bevat. Aan de ene kant accepteert Chalmers terecht dat we weten dat we fenomenaal bewust zijn, maar aan de andere kant volgt uit Chalmers’ theorie dat we dat niet kunnen weten. Vanzelfsprekend ontkent Chalmers dat dat laatste volgt en dat zijn theorie daarom met een probleem zou zitten.

Chalmers onderscheidt logische van natuurlijke superveniëntie, waarbij hij claimt dat fenomenaal bewustzijn wel natuurlijk maar niet logisch supervenieert: fenomenale ervaringen superveniëren _louter natuurlijk_. Het probleem zit hem in de notie van _louter natuurlijke superveniëntie_. Ik begin met het uiteenzetten van deze noties van superveniëntie.

_Logische superveniëntie:_ B-eigenschappen superveniëren logisch op A-eigenschappen, wanneer we kunnen zeggen dat de A-eigenschappen de B-eigenschappen insluiten, waarbij een feit een ander feit insluit als het logisch onmogelijk is voor de eerste om overeind te blijven zonder de tweede. (1996b: 36)

_Natuurlijke superveniëntie:_ B-eigenschappen superveniëren natuurlijk op A-eigenschappen als twee natuurlijk mogelijke situaties met dezelfde A-eigenschappen dezelfde Beigenschappen hebben. (1996b: 36)
Logische en natuurlijke superveniëntie vallen vaak samen. Chalmers wil echter laten zien dat dit voor fenomenaal bewustzijn niet geldt: fenomenaal bewustzijn is wel natuurlijk maar niet logisch superveniënt. Natuurlijke superveniëntie zonder logische superveniëntie is *louter natuurlijke superveniëntie*.

Omdat we weten waar Chalmers heen wil, is fenomenaal bewustzijn het beste voorbeeld om een test te maken voor loutere natuurlijke superveniëntie. Chalmers meent dat zowel zombies als omgekeerde tweelingen mogelijk zijn. Dat impliceert dat gegeven de A-feiten een B-eigenschap afwezig kan zijn (in het geval van de zombies), of anders (in het geval van de omgekeerde tweelingen). De volgende test volgt.

(T) **B-eigenschappen superveniëren louter natuurlijk op A-eigenschappen alleen dan als** voor twee werelden \( w_1 \) en \( w_2 \) die identiek zijn met betrekking tot de A-eigenschappen, het mogelijk is dat de B-eigenschappen die aanwezig zijn in \( w_1 \), anders of afwezig zijn in \( w_2 \). Met andere woorden, louter natuurlijk superveniërende eigenschappen kunnen onafhankelijk van hun A-eigenschappen variëren tussen werelden.

Deze test maakt duidelijk dat er geen eigenschappen zijn die tot het breed fysische domein behoren die louter natuurlijk superveniënt zijn, tenzij de fysische wetten niet tot de A-feiten behoren. Chalmers is ambigu over de status van de fysische wetten: al naar gelang het hem uitkomt behoren ze tot de A-feiten of niet. Mijn voorstel is dat hij de ambiguïteit als volgt oplost. Alle fysische wetten behoren tot de A-feiten, omdat anders er geen enkele logisch superveniënte eigenschap is: alle superveniërende eigenschappen zijn dan louter natuurlijk superveniërende eigenschappen. De vraag die we vervolgens kunnen stellen is: geef eens een voorbeeld van een louter natuurlijk superveniërende eigenschap. Dit moet dus een eigenschap zijn die buiten het breed fysische domein valt. Welk voorbeeld men ook geeft, het probleem hiermee is dat we er geen epistemologische relatie mee kunnen hebben, omdat een kennis toestand volgens Chalmers een breed fysische toestand is (in mijn termen). Dit moet dus de A-feiten veranderen ook in de wereld waarin de betreffende eigenschap niet voorkomt. Maar dan heeft men in die wereld kennis van iets wat niet bestaat, en kan er dus geen sprake zijn van een epistemische relatie. Ik ben van mening dat we enkel kennis kunnen verwerven via een causale relatie met het gekende, iets wat Chalmers in ieder geval voor fenomenaal bewustzijn ontkent. De vraag die nu dus volgt is: is fenomenaal bewustzijn wellicht speciaal in die zin dat het zowel louter natuurlijk superveniënt is, alsook kenbaar zonder causale relatie tussen de kennis toestand en de fenomenale ervaringen? Dat is precies wat Chalmers beweert. Het grote probleem is nu dat Chalmers de mysterieuze

182
epistemische relatie enkel een naam geeft ("acquaintance") maar dat hij die geenszins uitlegt. Om als oplossing voor het bovenstaande probleem te kunnen gelden, moet Chalmers deze bekendheid relatie uitleggen. Zolang hij dat niet doet, volgt uit de notie van louter natuurlijke superveniëntie dat dit soort eigenschappen – inclusief fenomenaal bewustzijn – onkenbaar zijn. Dit is contradictoir met de juiste aanneming dat we weten dat we fenomenaal bewust zijn. Chalmers’ theorie is dus incorrect. (Ik bespreek in hoofdstuk vier verder nog Chalmers’ retorische truc om de aandacht van dit probleem af te leiden, en laat zien dat hij daarmee hetzelfde probleem terug krijgt.)

HOOFDSTUK VIJF
GEEN TE STERKE EISEN

Hoofdstuk vijf is een relatief kort hoofdstuk waarin ik twee verwante argumenten bespreek: het vleermuis-argument van Thomas Nagel (1974) en het kennis-argument van Frank Jackson (1982). Nagel trekt uit zijn argument de conclusie dat we niet inzien hoe fysicalisme juist kan zijn, Jackson meent dat we kunnen concluderen dat het een onjuiste positie is. In beide gevallen volgt dat we geen wetenschappelijke theorie over fenomenaal bewustzijn op kunnen stellen.

In het kort komt mijn kritiek op beide argumenten op het volgende neer. Zowel Nagel als Jackson eisen van een theorie dat deze kan vertellen aan een blinde wat het is om rood te zien. Dit is een veel te sterke eis, analoog aan de eis dat een theorie over vloeibaarheid vloeibaarheid constitueert. Als je deze eis laat vallen, dan vervallen ook de argumenten van Nagel en Jackson die erop gericht zijn aan te tonen dat een fysicalistische theorie niet kan vertellen aan een blinde wat het is om rood te zien.

HOOFDSTUK ZES
DE NATUURLIJKE METHODE

Colin McGinn verdedigt een naturalistische en monistische positie. Desondanks meent hij dat we geen wetenschappelijke oplossing voor het lichaam-geest probleem kunnen vinden en wel om de reden dat we daar cognitief te beperkt voor zijn: we hebben de concepten er niet voor. Wat er zou moeten gebeuren is dat we op de een of andere manier de concepten die betrekking hebben op bewustzijn relateren aan die concepten die betrekking hebben op de fysische (neurofysiologische) wereld. Dit zal niet lukken, vanwege de homogeniteits-restrictie: om menselijk gedrag en handelen te kunnen verklaren kunnen we met onze fysische concepten toe: er is geen reden om concepten met betrekking tot bewustzijn te introduceren. Aan de andere kant
weten we via introspectie dat we bewuste wezens zijn. Hoe we het ook proberen: introspectie laat nooit zien dat voor bewustzijn een fysisch lichaam nodig is, en neurologisch onderzoek brengt ons nooit bij bewustzijn.

Owen Flanagan meent dat McGinn een methode over het hoofd ziet: de combinatie van introspectie met psychologie en neurowetenschap. Dit is wat hij de *natuurlijke methode* noemt.


In de hoofdstukken zeven tot en met elf wordt deze methode toegepast op de volgende typen fenomenaal bewustzijn: (1) fantoom ervaringen; (2) synesthetische ervaringen; (3) dromen; (4) kleurervaringen; (5) ervaringen van schoonheid van gezichten van soortgenoten.

In de samenvatting van deze hoofdstukken beperk ik me tot (1) een beschrijving van de betreffende fenomenale ervaringen; (2) de conclusies die uiteindelijk volgen uit de toepassing van de natuurlijke methode op de data die betrekking hebben op de betreffende fenomenale ervaringen.

**HOOFDSTUK ZEVEN**

**STUDIE # 1: ERVARINGEN VAN FANTOOMLEDEMATEN**

Sommige mensen hebben het ongeluk dat zij een of meerdere ledematen moeten missen. Vaak hebben deze mensen de ervaring alsof zij hun geamputeerde ledematen nog hebben. Dit zijn fantoomervaringen. Veelal zijn de ervaringen pijnlijk en gaat het dus om pijn die men lijkt te hebben in een deel van het lichaam wat er niet langer is. Dit maakt de pijn niet minder werkelijk.

De toepassing van de natuurlijke methode betreft in deze case studie (1) de fenomenologie combineren met (2) de neurologie van fantoomervaringen. Een belangrijke rol is hierbij weggelegd voor de representatie van het lichaam in de hersenen.

De gegevens over fantoomledematen laten het volgende zien. Fenomenale ervaringen kunnen aanwezig zijn in de afwezigheid van de normale fysische oorzaak. Als de representatie van het
lichaam in de hersenen op de juiste manier gestimuleerd wordt, dan resulteert dat in de beleving van een fenomenale wereld die buiten het hoofd gelokaliseerd lijkt te zijn. Echter, de ervaring moet in het hoofd gelokaliseerd zijn (vanwege de volgende redenering: (a) fenomenale ervaringen bestaan; (b) bestaan impliceert fysisch gerealiseerd zijn (of fundamenteel zijn, maar dat is bewustzijn niet); (c) als iets fysisch gerealiseerd is, dan heeft het ipso facto een locatie; (d) ergo fenomenale ervaringen moeten een locatie hebben; (e) omdat er geen realisatiebasis buiten het hoofd is voor fantoombervaringen moeten ze dus in het hoofd zitten. Omdat de ervaring in het hoofd zit en erbuiten lijkt te zijn, moet deze dus geprojecteerd worden naar de locatie waar deze lijkt te zijn.

Dit projectionisme lijkt in eerste instantie sterk op dat van Max Velmans (2000) maar verschilt voornamelijk met betrekking tot punt vier: Velmans meent dat de fenomenale wereld wel degelijk daarbuiten is, en niet in het hoofd.

HOOFDSTUK ACHT
STUDIE # 2: SYNESTHESIE

Synesthesie wordt door Simon Baron-Cohen en John Harrison (1997a) als volgt gedefinieerd:

> Wij, en anderen [...] definiëren synesthesie als optredend indien stimulering van een sensorische module automatisch een perceptie teweegbrengt in een tweede module, in de afwezigheid van enige directe stimulering van die tweede module.” (1997a: 3)

In hoofdstuk acht laat ik allereerst zien dat er vele varianten zijn van synesthetische ervaringen. Men kan een extra ervaring hebben van kleur als men iets proeft, hoort, voelt of ruikt, maar ook als men een bepaald cijfer ziet of er zelfs maar aan denkt. Niet alle combinaties zijn echter gerapporteerd: visuele input die resulteert in het horen van geluiden bijvoorbeeld lijkt niet voor te komen.

De toepassing van de natuurlijke methode bestaat uit de combinatie van (1) de fenomenologie, (2) de psychologie en (3) de neurologie van synesthesie.

De gegevens over synesthesie laten het volgende zien. Naast de vele varianten van synesthesie kunnen we synesthetische ervaringen indelen in twee soorten: (a) sensorische en (b) conceptuele synesthesie. Bij de eerste vormen normale sensorische stimuli de input die leidt tot de synesthetische ervaring, bij de tweede zijn dat concepten.
Door gebruik te maken van de conceptuele synesthesie kunnen we zien dat externalisme met betrekking tot fenomenale inhoud niet houdbaar is. (Voor sensorische synesthesie geldt hetzelfde, maar bij de conceptuele soort is de conclusie veel helderder.) Het argument hiervoor gaat als volgt: (a) het fenomenale is de representationele inhoud; (b) een synestheet heeft twee ervaringen bij één input, en daarmee dus extra inhoud; (c) deze extra inhoud kan enkel internalistisch verklaard worden: (c₁) als we de extra ervaring zien als een representatie, dan moeten we zeggen – bij conceptuele synesthesie – dat de extra ervaring een concept representeert; (c₂) een concept is zelf een representatie, en dus is de extra ervaring een representatie van een representatie; (c₃) externalisme sluit secundaire eigenschappen uit (McGinn 1989); (c₄) omdat concepten niet gekleurd zijn is het zeker dat in dit geval de kleuren secundaire eigenschappen zijn van het concept en dus kan een externalistische verklaring hier niet juist zijn.

De neurologie van conceptuele synesthesie laat zien dat de activering van het neurale correlaat van het concept de oorzaak is van de activering van het neurale correlaat van de extra ervaring. Omdat dit een ongebruikelijke manier is voor het veroorzaken van fenomenale ervaringen (het lijkt een gevolg te zijn van het niet afsluiten van neurale paden tussen de desbetreffende modules) kunnen we zeggen dat het hier in eerste instantie gaat om een ervaring die niet bedoeld is om te representeren: de synesthetische ervaringen zijn allereerst louter presentaties. Het is duidelijk dat deze presentaties gebruikt kunnen worden als representaties: een synestheet kan bijvoorbeeld blauw gebruiken om het concept vier te representeren.

Een extra conclusie van (c₄) is dat fenomenale ervaringen hogere-orde representaties kunnen zijn, waarmee een eerste-orde theorie over fenomenale ervaringen dus principieel tekort schiet.

De objectie die men kan maken en die ervoor zorgt dat geen van de bovenstaande conclusies zou volgen is: synesthetische ervaringen zijn misrepresentaties. Het argument in het kort tegen deze opvatting van synesthetische ervaringen is het volgende. Neem bijvoorbeeld een geluid dat ook als rood ervaren wordt. Het geluid wordt niet verkeerd gerepresenteerd als het gaat om de representatie van het geluid als een geluid. Van totale misrepresentatie (zoals een kat die we voor een konijn aanzien) is dus geen sprake. Het moet dus gaan om de misrepresentatie van een eigenschap van het geluid. Maar welke eigenschap is dat dan? Het antwoord is dat die eigenschap er niet is, waardoor aan een belangrijke eis voor misrepresentatie niet voldaan is: er moet iets zijn dat verkeerd gerepresenteerd wordt. Bovendien zijn de extra ervaringen consistent doorheen de
tijd, waardoor ze gebruikt kunnen worden als betrouwbare representaties, iets wat met misrepresentaties niet het geval is.

De conclusies met betrekking tot synesthetische ervaringen staan dus overeind. Deze conclusies suggereren het volgende.

(1) In het geval van synesthesie is het duidelijk dat externalisme onjuist is en dat kleuren secundaire eigenschappen zijn. Wellicht geldt dit voor alle fenomenale ervaringen.

(2) Synestheten gebruiken hun extra presentaties als representaties. Wellicht hebben alle mensen de natuurlijke neiging om presentaties onmiddellijk als representaties te interpreteren, ook als ze dat niet zijn.

In hoofdstuk negen laat ik aan de hand van dromen zien dat de tweede suggestie klopt. Ook hier volgt weer de eerste suggestie. In hoofdstuk tien toon ik aan dat ook die juist is.

**HOOFDSTUK NEGEN**

**STUDIE # 3: DROMEN**

Hoofdstuk negen is voor een groot deel gebaseerd op het broek van Flanagan *Dreaming Souls* (2000) waarin hij de natuurlijke methode toepast op dromen. In dit hoofdstuk laat ik zien dat de fenomenale herrie die tijdens de slaap gegenereerd wordt als bijproduct van slapen geïnterpreteerd wordt als representaties. Als verdere conclusie volgt dat fenomenale ervaringen in het hoofd gelokaliseerd zijn.

De fenomenologie van dromen vertelt ons onder andere dat dromen in twee soorten komen. Ten eerste de bizarre dromen die erg levendig zijn en voornamelijk visueel zijn. Ten tweede de dromen die bestaan uit een klein probleem (bijvoorbeeld de vraag waar je je schoenen gelaten hebt) en wat alsnog herhaald wordt. Om de een of andere reden is dit klein probleem in je slaap een erg groot probleem, en deze soort dromen heeft dan ook het karakter van een nachtmerrie. Verder is het zo dat we dromen erg slecht of helemaal niet onthouden. Dit suggereert dat het goed mogelijk is dat we de hele nacht onafgebroken dromen.

De empirische data die ik in hoofdstuk negen presenteer laten zien dat we zo slecht zijn in het onthouden van dromen omdat ‘s nachts de neurotransmitters die nodig zijn voor het onthouden van zaken niet afgescheiden worden door de neuronen die ze aanmaken.
De vraag of we wellicht de hele nacht dromen, en dus de hele nacht bewust zijn kunnen we ook positie beantwoorden. De these dat voor bewustzijn 40 Hz oscillaties in de hersenen nodig zijn is inmiddels erg goed gecorroboreerd, zodat we als we deze oscillaties aantreffen in de hersenen we kunnen concluderen dat de persoon in kwestie bewust moet zijn. Tijdens alle fasen van dromen treffen we deze oscillaties aan.

De evolutie-biologie van dromen laat zien dat 's nachts de hersenen actief zijn met het herstellen van de niveaus van neurotransmitters en hormonen die overdag nodig zijn. Tijdens deze periode wordt één functie echter niet uitgeschakeld, namelijk die van het interpreteren van de presentaties als representaties: dit maakt dromen tot bijverschijnselen van het interpreteren van presentaties die door reguliere stimuli veroorzaakt worden. Omdat het evident is dat deze reguliere stimuli tijdens de slaap niet de oorzaak van de presentaties zijn, is het ook duidelijk dat een externalistische kijk op deze fenomenale ervaringen niet juist kan zijn: dromen plaatsten fenomenale (re)presentaties stevig in het hoofd.

Omdat internalisme hand in hand gaat met een secundaire kijk op – onder andere – kleuren, en die suggestie ook voortkwam uit het toepassen van de natuurlijke methode op synesthesie, kijk ik in hoofdstuk tien naar kleuren.

HOOFDSTUK TIEN
STUDIE # 4: KLEURERVARINGEN

Indien kleuren secundaire eigenschappen zijn, dan zijn ze daarmee voor hun bestaan afhankelijk van (1) een waarnemer en (2) de primaire eigenschappen van het object waarvan we zeggen dat het een kleur heeft. Dit biedt twee manieren om te laten zien dat kleuren inderdaad secundaire kwaliteiten zijn. Ten eerste kunnen we laten zien dat kleuren afhankelijk zijn van de waarnemer, door naar de fenomenologie van verschillende waarnemers te kijken. Ten tweede kunnen we aantonen dat er geen primaire eigenschappen zijn die in aanmerking komen om geïdentificeerd the worden met kleuren. In beide gevallen heeft het argument de volgende structuur: (a) wat is het voorstel van de verdedigers van de claim dat kleuren primaire eigenschappen zijn? Is dat identificatie met een eigenschap van licht of van oppervlakten van objecten; (b) laat in beide gevallen zien dat de voorstellen neerkomen op het accepteren van bijvoorbeeld rood als een disjunctieve eigenschap; (c) laat vervolgens zien dat een excessief disjunctieve eigenschap is,
met andere woorden: laat zien dat het dan om meerdere verschillende eigenschappen gaat; (d) als dat zo is dan kan je vervolgens laten zien dat er slechts één de echte kleur rood is, en de andere zijn niet echt. Maar die laatste claim impliceert dus dat wij iets dat niet rood is ervaren als iets dat wel rood is. Maar dat maakt dus van deze gevallen van namaak-rood secundaire eigenschappen.

Door naar kleurenblindheid te kijken komen we achter een manier om het eerste punt te bewijzen: kleuren zijn afhankelijk van waarnemers. De meerderheid van de mensen heeft wat we normale kleurenvisie noemen. Sommige mensen – voornamelijk mannen – zijn “kleurenblind”: ze ervaren wel alle kleuren maar vaak verwarren ze verschillende kleuren. Stel je nu voor dat er mensen zijn die echter nog meer kleuren onderscheiden dan mensen met normale kleurenvisie, dan zijn voor die mensen diegenen die we gewoonlijk als niet-kleurenblind beschouwen kleurenblind. De empirische data laten zien dat een zeer klein percentage van de vrouwen waarschijnlijk extra kleuren ervaart ten opzichte van de mensen met normale kleurenvisie. Het is daarmee duidelijk dat de kleur die een object heeft afhankelijk is van de waarnemer en dus dat kleuren secundaire eigenschappen zijn.

Een tweede mogelijkheid om te laten zien dat kleuren secundaire eigenschappen zijn, is om te laten zien dat het geen primaire eigenschappen kunnen zijn: er is niet één primaire eigenschap die bijvoorbeeld rode objecten met elkaar gemeenschappelijk hebben, maar er zijn vele van dit soort eigenschappen, die onderling zo verschillend zijn dat ze niet als één eigenschap kunnen beschouwen. Als we kijken naar wat de voorstanders van de claim dat kleuren primaire eigenschappen aanwijzen als wat dat dan zou zijn, dan is er een tweetal mogelijkheden: (1) het is een (disjunctieve) eigenschap van het licht; (2) het is een (disjunctieve) eigenschap van het oppervlak van een gekleurd object.

Kunnen we bijvoorbeeld de kleur geel identificeren met een bepaalde samenstelling van licht? Het fenomeen kleurconstantheid geeft aan dat dat niet kan: een banaan is geel of deze nu verlicht wordt door daglicht, kaarslicht, in de schemer of als er een lamp op schijnt. In deze gevallen is de samenstelling van het licht zo verschillend dat we moeten spreken van een disjunctieve eigenschap. Mijns inziens is de disjunctie hier al zo breed dat we het eigenlijk niet meer over één disjunctieve eigenschap kunnen hebben, maar dat dit een excessief disjunctieve eigenschap is. Jackson stelt echter voor om dit toch als kandidaat voor kleur te beschouwen. Maar dan volgt eveneens dat de disjunctieve eigenschap excessief disjunctief is. Het is experimenteel vastgesteld dat eenzelfde samenstelling van licht verschillende kleurervaringen teweeg kan brengen.
Samenstelling A veroorzaakt bijvoorbeeld rood en groen, en samenstelling B rood en bruin. Dat impliceert dus dat zowel samenstelling A als samenstelling B in de disjunctieve eigenschap van rood terecht komt, wat deze eigenschap dermate excessief maakt dat we niet langer meer over één eigenschap kunnen spreken. Daarbij kunnen we dan niet langer verschil maken tussen rood, groen en bruin, omdat deze allemaal dezelfde eigenschap in hun disjunctieve eigenschap hebben.

De poging om kleuren te identificeren met eigenschappen van objecten komt ook uit bij het accepteren van excessief disjunctieve eigenschappen. Het probleem is nu dat experimenteel is vastgesteld dat als we waarnemer, samenstelling van het licht en object constant houden, en enkel de condities van de omgeving van het object veranderen, het object van kleur verandert. De betreffende primaire eigenschappen van het oppervlak van het object kunnen dus eenvoudigweg niet met één bepaalde kleur geïdentificeerd worden, omdat dat dan ook moet gebeuren met een andere kleur. Dit houdt verder (alweer) in dat als we kleuren willen identificeren met een disjunctie van eigenschappen het zo zal zijn dat een eigenschap die bijvoorbeeld in de disjunctieve eigenschap rood komt te zitten, ook in de disjunctieve eigenschap grijs komt te zitten, waardoor er geen verschil meer is te maken tussen rood en grijs (in ieder geval niet op het niveau van primaire eigenschappen). Met andere woorden: de disjunctie wordt excessief, waardoor we dus over verschillende eigenschappen praten.

Jackson meent dat er een oplossing bestaat: kleuren zijn excessief disjunctieve eigenschappen. Met andere woorden: er zijn meerdere soorten rood. Jackson meent echter dat er nu sprake is van echt rood en van nep-rood. Maar het accepteren van nep-rood is eenvoudigweg stellen dat er een primaire eigenschap is die (1) niet echt rood is, maar die (2) wel rood-ervaringen in ons veroorzaakt. Nep-rood is dus een secundaire eigenschap van een object dat wel rood lijkt maar het niet is. Dat is precies wat de theorie die de excessief disjunctieve eigenschappen moet accepteren afwijst, waardoor de theorie zichzelf weerlegd.

**HOOFDSTUK ELF
STUDIE #5: ERVARINGEN VAN SCHOONHEID**

Mensen hebben ervaringen van schoonheid bij gezichten. In tegenstelling tot wat de meeste mensen denken zijn deze ervaringen intersubjectief: we ervaren dezelfde gezichten als mooi. Ten eerste zijn gezichten die gemiddeld zijn erg mooi. Ten tweede wijken de gezichten die we het mooist vinden allemaal op een specifieke manier af van dat gemiddelde. Zo zijn de mooiste gezichten van vrouwen gezichten waarbij de lippen net iets voller zijn en de onderkaak iets
kleiner is. De mooiste mannengezichten hebben precies een grotere onderkaak dan het gemiddelde mannengezicht (indien geëvalueerd door een vrouw die in haar vruchtbare periode is). Deze gezichtskenmerken komen overeen met een bepaalde mix van testosteron en oestrogeen in de pubertijd. Veel testosteron en weinig oestrogeen leidt tot een extreme groei van de onderkaak. De tegenovergestelde mix leidt tot een remming van die groei. De mix van weinig testosteron en veel oestrogeen zorgt er bij vrouwen ook voor dat zij erg vruchtbare zijn. De tegenovergestelde mix bij mannen is een indicatie van hun goede genotypische en fenotypische gezondheid.

Zonder dat we het zelf weten, gebruiken we dus ervaringen van schoonheid om de meest vruchtbare vrouwen en de gezondste mannen te herkennen. Deze manier van herkennen is direct. Het gaat immers om een presentatie die gebruikt wordt als een representatie en presentaties zijn altijd direct. Daarnaast is het een indirecte representatie, omdat de vruchtbaarheid van de vrouw niet direct geregistreerd wordt, maar enkel de secundaire eigenschap aantrekkelijk zijn. Het gaat daarmee om een zeer economische manier van het representeren van de vruchtbaarheid van een vrouw (of de gezondheid van een man) zonder dat we kennis vereist is van de neodarwiniaanse evolutie-theorie. Precies omdat er zeer snel, kort en indirect evolutionair belangrijke en complexe informatie wordt geregistreerd noem ik fenomenaal bewustzijn het steno van de evolutie.

**HOOFDSTUK TWALF**

**DRIE BEDREIGINGEN VOOR REPRESENTATIONALISME GENEUTRALISEERD**

Nu de evolutionair steno theorie van de fenomenale geest ondersteund is met empirische data, rest het me nog om een drietal argumenten tegen representationele theorieën van fenomenaal bewustzijn van repliek te dienen.

Een positieve kijk op representationele theorieën van bewustzijn is de volgende. We begrijpen hoe representaties (zoals foto’s) werken en als we nu van bewuste ervaringen kunnen laten zien dat ze ook representaties hebben, dan ligt er een verklaring klaar voor bewuste ervaringen. Uit deze positieve kijk volgt een algemeen kritiekpunt op representationele theorieën van bewustzijn. In het algemeen zijn representationele toestanden niet bewust, dus theorie van bewustzijn laat precies bewustzijn weg als men bewuste ervaringen reduceert tot representaties. Een representationele theorie loopt dus het risico bewustzijn te verliezen.
Er is een aantal redenen waarom de evolutionair steno theorie niet het risico loopt fenomenale ervaringen te reduceren tot representaties. Ten eerste claimt de theorie enkel dat bewuste ervaringen gebruikt kunnen worden als representaties, niet dat het altijd representaties zijn. Reductie is daarmee onmogelijk voor die representaties die niet als representatie gebruikt worden.

Ten tweede creëren fenomenale representaties intensionele toestanden waardoor we geen niet-fenomenale representaties voor ze niet mogen substitueren, en ze daar dus niet toe kunnen reduceren. Een ervaring van schoonheid is en blijft een ervaring van schoonheid, ook al representeren het indirect de vruchtbaarheid van degene waarnaar je kijkt. De meeste mensen realiseren zich niet eens dat dat zo is. Een reductie van ervaringen van schoonheid naar kennisstochten met betrekking tot de vruchtbaarheid van de persoon die men ziet is dan ook uitgesloten.

Een tweede punt van kritiek op een representationele theorie van bewustzijn komt voort uit de volgende redenering: fenomenale toestanden worden bepaald door neurale toestanden. Als die neurale toestanden iets representeren, dan is het plausibel dat de fenomenale toestanden hetzelfde representeren. Dit is de positie van Max Velmans. Dit leidt tot wat ik het probleem van overrepresentatie noem: alle representationele werk wordt al gedaan door de neurale representaties, waardoor de fenomenale representaties overbodig worden. (Dit probleem is verwant aan dat van causale overdeterminatie.) Ook hier dreigt weer reductionisme, in dit geval van de fenomenale naar de neurale representaties.

Het argument tegen de veronderstelling dat de neurale toestand hetzelfde representeert als de fenomenale toestand die erdoor bepaald, komt van de studie naar ervaringen van schoonheid. De directe representatie van een mooi gezicht is dat van het gezicht als zijnde mooi (het gaat hier nu niet om de indirecte representatie van vruchtbaarheid of gezondheid). Het is duidelijk dat op fenomenaal niveau de verhoudingen tussen de ogen, neus, mond etcetera niet gedetailleerd gerepresenteerd worden. (Was dat wel het geval, dan had men geen onderzoek hoeven doen naar wat die verhoudingen zijn voor mooie gezichten.) Maar op het – onbewuste – niveau van de neurale representaties zal dat wel nodig zijn, omdat het precies die verhoudingen zijn die de gevoelens van schoonheid bepalen. Wat we hier zien is dat de neurale representatie van een gezicht een heel andere inhoud heeft dan de fenomenale representatie die erdoor bepaald wordt. Het is daarmee niet langer vol te houden dat beide representaties hetzelfde representeren waardoor de fenomenale representatie een representationeel bijverschijnsel zou worden.
Een derde objectie tegen een representationele theorie van de fenomenale geest is dat bewustzijn überhaupt niets met representaties te maken heeft. Kevin O’Regan en Alva Noë (2001) verdedigen zo’n theorie. Zij zijn van mening dat het voldoende is voor een bepaald organisme om de wetten te kennen die betrekking hebben op (1) de veranderingen die de dingen veroorzaken die je waarneemt (bijvoorbeeld het geluid dat een auto maakt als die voorbij rijdt) en (2) het waarnemingsapparaat (bijvoorbeeld het oor). Als een organisme de betreffende wetten beheerst, dan heeft het geen representaties nodig om zich in de wereld te bewegen. De wereld dient zelf dan als een representatie van zichzelf.

Ik meen dat hier drie grote problemen mee zijn. Ten eerste geven O’Regan en Noë zelf aan dat hun theorie niet enkel opgaat voor bewuste organismen, maar ook voor een raketgeleidingssysteem. Hierdoor is het onduidelijk waarom dit een theorie over bewustzijn zou zijn.

Ten tweede kan de theorie niet uit de voeten met keuzes maken en planning. Wil men niet enkel direct reageren op bepaalde stimuli, maar wil men keuzen kunnen maken die de omgeving veranderen, dan zal men de omgeving op verschillende manieren voor moeten kunnen stellen, om zo te kunnen bepalen welke verandering aan de omgeving de meest wenselijke is. Dat is dus enkel mogelijk als men de omgeving op verschillende manieren kan representeren.

Ten derde, volgens O’Regan en Noë is representationalisme gecommitteerd aan de claim dat we een gedetailleerde representatie hebben van datgene waar we naar kijken. Ze geven overtuigend empirisch materiaal om aan te tonen dat dat niet zo is: we missen zelfs een vrouw in een gorillapak die door een veldje van zes basketballende mensen loopt. In plaats van te stellen dat de representatie niet gedetailleerd is, trekken O’Regan en Noë de conclusie dat representationalisme geen correcte theorie is. Mijns inziens moet men de eerste conclusie trekken. Deze is ook veel plausibeler vanuit een evolutionair gezichtspunt: het is veel economischer om enkel de belangrijke zaken bewust te representeren dan om alles te representeren. Het zou dan veel moeilijker worden om beslissingen te nemen, omdat men eerst nog uit moet zoeken wat wel en wat niet belangrijk is.

Met dit weerwoord is ook de laatste van de drie objecties tegen representationalistische theorieën van fenomenaal bewustzijn weerlegd. In ieder geval kan men ze niet langer in stelling brengen tegen de evolutionair steno theorie van de fenomenale geest.
REFERENCES


Hitt, J. (1999), "This Is Your Brain on God, Michael Persinger has a vision – the Almighty isn't dead, he's an energy field. And your mind is an electromagnetic map to your soul," Wired(7)(11): 308-314.

Jackson, F. (1982), "Epiphenomenal Qualia"


CURRICULUM VITAE

- 30 mei 1985: Diploma HAVO; Dr. Schaepmancollege (Dongen).
- 13 juni 1988: diploma VWO; Theresialyceum (Tilburg).

18 augustus 1997: diploma KUB
titel scriptie: Metaforen voor de geest – waarom we enkel in metaforen over de geest kunnen denken.
afgestudeerd bij: dr. F. Buekens, Prof. dr.A. Derksen & Prof. Dr. H. de Swart.
werkzaamheden tijdens deze studie:
augustus 1995 – juli 1997: student-assistent van dr. F. Buekens, Prof. Dr. H. de Swart & Prof. Dr. G. Berns.