

UNIVERSITY OF STIRLING

Department of Philosophy

The Knowledge Argument

LUCA MALATESTI

Thesis submitted for the degree of Doctor of Philosophy

March 2004

For Nela

Abstract

Frank Jackson's knowledge argument is a very influential piece of reasoning that seeks to show that colour experiences constitute an insoluble problem for science. This argument is based on a thought experiment concerning Mary. She is a vision scientist who has complete scientific knowledge of colours and colour vision but has never had colour experiences. According to Jackson, upon seeing coloured objects, Mary acquires new knowledge that escapes her complete scientific knowledge. He concludes that there are facts concerning colour experiences that scientific knowledge can neither describe nor explain. Specifically, these facts involve the occurrence of certain non-physical properties of experiences that he calls *qualia*.

The present research considers whether a *plausible formulation* of the hypothesis that science can accommodate colour experiences is threatened by a *version* of the knowledge argument. The specific formulation of this problem has two motivations. Firstly, before investigating whether the knowledge argument raises a problem for the claim that science can account for colour experiences, we need a plausible formulation of this claim. I argue that the idea that science can accommodate colour experiences can be formulated as the *modest reductionism hypothesis*. Roughly speaking, this is the hypothesis that a science that can be explanatory interfaced with current physics of ordinary matter can account for conscious experiences.

Secondly, an unintelligible premise figures in Jackson's version the knowledge argument. Namely, it is assumed that Mary possesses a complete (future or possible) scientific knowledge. Nevertheless, the type of strategy involved in Jackson's argument can be used to target modest reductionism. By considering contemporary psychophysics and neuroscience, I characterise Mary's scientific knowledge. First, this characterisation is intelligible. In fact, it is elaborated on the basis of descriptions and explanations of colour experiences involved in current physics and neuroscience. Second, a supporter of modest reductionism can assume that the scientific knowledge ascribed to Mary might account for colour experiences.

The main conclusion of the present research is that our version of the knowledge argument fails to threaten the modest reductionism hypothesis. In fact, I endorse what can be called the "two ways of thinking" reply to the knowledge argument. According to this response, the knowledge argument shows that there are different ways of thinking about colour experiences. One way of thinking is provided by scientific knowledge. The other way of thinking is provided by our ordinary conception of colour experiences. However, the existence of these two ways of thinking does not imply the existence of facts and properties that escape scientific knowledge. It might be the case that the ordinary way of thinking about colour experience concerns facts and properties described and explained by science.

The principal conclusion of the research results from two investigations. The first line of research aims to reveal and evaluate the implicit assumptions that figure in the knowledge argument. The main body of the research is dedicated to this task. The principal result of this investigation is that the knowledge argument must rely on an account of introspective knowledge of colour experiences. I argue that an inferential model of introspection provides such account. On this model, Mary's capacity to hold beliefs about her colour experiences when she sees coloured objects requires her mastery of colour concepts. The second main investigation seeks to justify the two ways of thinking strategy. As many opponents and supporters have recently started to realise, this strategy might be charged with being *ad hoc*. I offer a distinctive justification of this reply to the knowledge argument. Assuming the account of introspection mentioned above, the existence of *visual recognitional colour concepts* might justify this strategy. A person possesses these concepts when she is able to determine the colours of objects simply by having visual experiences.

Contents

Preface	vi
Introduction	1
1 The Main Problem and its Solution	1
2 The Implicit Premises of the Knowledge Argument	3
3 Two Ways of Thinking About Colour Experiences	4
4 The Plan of the Volume	6
1 A Hypothesis about Conscious Experiences	9
1.1 Introduction	9
1.2 Defining “Being Physical”	10
1.3 The Limitations of Classical Reductionism	25
1.4 Modest Reductionism.....	37
1.5 Conclusion.....	39
2 The Knowledge Argument	41
2.1 Introduction	41
2.2 A Thought Experiment about Mary	42
2.3 Knowledge Beyond our Understanding	48
2.4 Revising the Knowledge Argument	54
2.5 Conclusion.....	56
3 Mary's Scientific Knowledge	58
3.1 Introduction	58
3.2 Colour Spaces.....	58
3.3 The Scientific Account of Colour Experiences.....	63
3.4 Mary's Complete Knowledge	69
3.5 Patricia Churchland and Daniel Dennett Revisited.....	74
3.6 Conclusion.....	79

4	Knowing Colour Experiences.....	80
4.1	Introduction	80
4.2	The Fundamental Question about Mary	80
4.3	A Problem Concerning the Content of Mary’s Belief.....	90
4.4	Direct Awareness of Experiences and Perception.....	96
4.5	Introspection and the Direct Awareness of Experiences.....	100
4.6	Conclusion.....	106
5	The Content of Mary’s Belief	107
5.1	Introduction	107
5.2	An Account of Introspection	108
5.3	Connecting Beliefs	114
5.4	Representationalism and Introspection	120
5.5	The Content of Mary’s Belief	125
5.6	Conclusion.....	130
6	Resisting the Ontological Conclusion	131
6.1	Introduction	131
6.2	The Ability Reply.....	133
6.3	A Version of the Ability Reply	136
6.4	Resisting the Ability Reply	141
6.5	The “Two Ways of Thinking” Reply	145
6.6	Requirements on Phenomenal Concepts	151
6.7	Conclusion.....	155
7	Different Ways of Thinking About Colour Experiences.....	157
7.1	Introduction	157
7.2	The Indexical Reply	158
7.3	Against the Indexical Reply	169
7.4	Recognitional Concepts.....	176
7.5	Old Thick Properties	181
7.6	Conclusion.....	185
	References	188

Preface

I wish to thank the supervisors of this thesis. Professor José Luis Bermúdez, who has followed most of this work, has always been extremely patient, dedicated and rigorous. Mr Toni Pitson, who supervised the work for a semester, aided my understanding of many details of contemporary colour science. Professor Alan Millar had a fundamental role in motivating and helping me in the latter stages of my research.

I am also grateful to Dr Fiona Macpherson and Professor Peter Sullivan for reading earlier drafts of the first chapters and providing detailed comments. Moreover, I would like to thank Dr Garry Young for his accurate proofreading.

Parts of this work descend from published material. Sections of Chapter 4 derive from my “Conoscenza dei *qualia* e punti di vista” (“Knowledge of *Qualia* and Points of View”).¹ Others parts of Chapter 4 and Chapter 5 figure in “Conoscere le esperienze dei colori” (“Knowing Colour Experiences”) (forthcoming).² Finally, Chapter 6 contains material that derives from “Knowing what it is Like and Knowing How”.³ I would like to thank the editors and the referees for publishing these papers and providing comments and suggestions.

I had the great opportunity to try out some ideas of this work in research seminars and conferences in Stirling, Hull, Siena, Arezzo, Florence, Bergamo and Genova. I am grateful to the organisers for inviting me or accepting my papers. In addition, I would like to thank those who made comments on these occasions.

¹ Published in R. Lanfredini, ed. *Mente e Corpo: La soggettività fra scienza e filosofia*. Milano: Guerini, 2003.

² To appear on line in *Networks, Refereed Online Journal, Monographic Issue on Consciousness*, edited by S. Gozzano, (forthcoming): <http://lgxserver.ciseca.uniba.it/lei/ai/networks/04-1/> .

³ Published in A. Peruzzi, ed. *Mind and Causation*. Amsterdam: Benjamin, 2004.

I owe a great debt of gratitude to the Department of Philosophy at the University of Stirling. Over the years it has been a very friendly, supportive, and stimulating environment. I am very grateful to the members of the academic staff, the other PhD students, the secretaries, and the students who attended the seminars I taught as a teaching assistant. From each of them, I have learned a lot about British culture and language, philosophy, and the importance and beauty of teaching it.

In addition, I would like to thank the staff of the Department of Philosophy at the University of Hull. In the last period of my research, they have kindly provided me a place in the *Graduate Research Institute* of the University. Moreover, they have involved me in their teaching and research activities.

I am very grateful to Professor Mariano Bianca, Dr Gideon Makin, Santiago Amaya, Jordi Sala, Dr Daniela Sime, Dr Fiona Chalamanda, for their friendship and encouragement.

I wish to thank my parents, my brother, my sister, and the rest of my family for their constant support. Moreover, I am immensely grateful to my fiancée Nela. Despite the hard demands of her own research, she has helped (and tolerated) me in these years in every possible way. This work is about a fictional scientist, but it exists thanks to an actual one.

I would like to acknowledge for funding the *Student Awards Agency for Scotland*, the Faculty of Arts and the Department of Philosophy at the University of Stirling.

Introduction

1 The Main Problem and its Solution

A wide debate in contemporary philosophy of mind focuses on a general problem concerning conscious experiences. Many philosophers maintain that these mental states can be completely described and explained in scientific terms. Usually, this claim is related to the ontological statement that mental states are physical states, often understood as brain states. However, others argue that conscious experiences pose a fundamental problem for scientific knowledge. Specifically, some authors attack the epistemological claim that scientific knowledge can accommodate these mental states.¹ According to these philosophers, when we perceive a yellow lemon or we endure a pain, certain properties are instantiated that can neither be described nor explained by science.

Frank Jackson has offered a very influential argument for the claim that conscious colour experiences constitute an insoluble problem for science.² This argument is based on a thought experiment concerning Mary. She is a vision scientist who has complete scientific knowledge of colours and colour vision and who has never had colour experiences. According to Jackson, upon seeing coloured objects, Mary acquires knowledge that escapes her complete scientific knowledge. He concludes that there are facts concerning colour experiences that are beyond the scope of her scientific knowledge. Specifically, these facts involve the occurrence of certain non-physical properties of experiences called *qualia*.

¹ Influential formulations of this criticism can be found in Chalmers 1996, McGinn 1991, Nagel 1974.

² Jackson 1982, Jackson 1986.

The present research investigates whether *a certain formulation* of the hypothesis that science can account for colour experiences is threatened by a *version* of Jackson's knowledge argument. There are two motivations for stating the problem in these terms. Firstly, let us clarify why we have to focus on a certain version of the claim that science can describe and explain colour experiences. In Chapter 1, I show that this claim needs to be formulated in a plausible and substantive manner. I suggest that the *modest reductionism hypothesis* offers such a formulation. Roughly, this is the hypothesis that a science, which can be explanatorily interfaced with current physics of ordinary matter, can account for conscious experiences.

Secondly, we have to consider a certain *version* of the knowledge argument. In fact, Jackson's argument involves the unintelligible premise that Mary has complete (future or possible) scientific knowledge.³ Therefore, we cannot judge the soundness of this line of reasoning. Nevertheless, the type of strategy offered by Jackson can still be used to target the hypothesis of modest reductionism. In Chapter 3, I characterise Mary's scientific knowledge. First, this characterisation is intelligible. In fact, it is elaborated on the basis of descriptions and explanations of colour experiences involved in current physics and neuroscience. Second, modest reductionists can plausibly assume that the scientific knowledge delineated by this characterisation can account for colour experiences.

The present research's main conclusion is that our version of the knowledge argument does not threaten the modest reductionism hypothesis. I endorse what can be called the "two ways of thinking reply".⁴ According to this response, the knowledge argument shows that there are different ways of thinking about colour

³ This criticism is advanced in Dennett 1991, pp. 399-403; Churchland 1986, pp. 331-334.

⁴ This reply is also called the "new mode of presentation reply", "conceptual dualism strategy", "new knowledge and old fact reply". Amongst the recent upholders of this strategy, see Peacocke 1989, Loar 1990, Papineau 2002, Sturgeon 2000, Carruthers 2000, Tye 2000, Perry 2001. Precursors of the kind of view involved in this reply can be found in Feigl 1960, Smart 1959, Place 1956.

experiences. One way is provided by scientific knowledge. The other way is offered by our ordinary conception of colour experiences. However, the knowledge argument is not successful in supporting its ontological conclusion. Such reasoning does not show that mental states have proprieties beyond the scope of scientific knowledge.

The conclusion that our version of the knowledge argument is unsound is reached in virtue of two investigations. The first line of investigation, carried forward in Chapters 2 - 5, seeks to reveal and evaluate the implicit assumptions that figure in the knowledge argument. The second main investigation, which starts in Chapter 6 and continues in Chapter 7, assesses the knowledge argument. Specifically, I justify the two “ways of thinking strategy”. The remainder of the present introduction outlines these two investigations.

2 The Implicit Premises of the Knowledge Argument

The main body of this volume is dedicated to the elucidation of the implicit assumption that figure in the knowledge argument. The vast literature on this argument contains important insights about the structure of the argument and the nature of its premises.⁵ By considering and evaluating these different suggestions, I offer a comprehensive analysis of the structure and premises of the knowledge argument.

The elucidation of the knowledge argument brings about an important result: namely, Mary, upon her release, can acquire new knowledge about colour experiences only if she acquires new knowledge about the colours that objects look to have to her. Let us see why this is the case.

⁵ Detailed analyses of the knowledge argument can be found in Alter 1995, Perry 2001, Lewis 1990, Churchland 1989 (especially the postscript).

According to a standard interpretation, Mary comes to know that experiences have *qualia*. This interpretation requires that, by seeing coloured objects, Mary can form beliefs about the type of colour experience she is having. Moreover, having colour experiences should enable her to have beliefs based on an ordinary classification of colour experiences. On this classification, colour experiences differ when they have different *qualia*. Thus, she can discover that a certain type of colour experience, as specified by her scientific knowledge, has a property that escapes scientific description and explanation.

The knowledge argument does not explain how Mary can form the new beliefs that colour experiences have *qualia*. Such an explanation requires accounting for the transition from seeing coloured objects to acquiring these beliefs. Chapter 4 and 5 consider how the upholder of the knowledge argument might offer this account. I deny that Mary's supposed new beliefs could be based on her direct awareness of colour experiences or their properties. Neither perception nor introspection can provide this awareness.

A more plausible account is that Mary forms these beliefs in virtue of certain other cognitive capacities. First, Mary must possess the capacity to have thoughts concerning the colours objects look to have to her. Second, she needs to know that certain relations hold between (i) having a colour experience of a certain type and (ii) the fact that something looks a certain colour to her. Therefore, establishing what she learns about colour experiences requires another investigation. Namely, we have to establish what she might learn about the colour of objects.

3 Two Ways of Thinking About Colour Experiences

Chapter 6 and 7 consider whether Mary, by acquiring new knowledge about colour experiences, comes to know facts that escape her scientific knowledge. Firstly, I show that an important objection against this ontological conclusion of the knowledge argument fails. Some have resisted this argument by means of the *ability*

reply. According to this strategy, upon seeing coloured objects, Mary simply acquires a new set of abilities. Nevertheless, she does not form any new belief that she could not have already formed in her room. Therefore, she cannot acquire new propositional knowledge about facts her scientific knowledge fails to consider and explain. I show that the promoter of the knowledge argument can resist this objection. However, conceding that Mary acquires new beliefs about colours does not imply that there are facts concerning colour experiences that escape her scientific knowledge.

I endorse a reply to the knowledge argument that is becoming a standard move amongst many philosophers. This reply, which I dubbed the “two ways of thinking” strategy, assumes the consistency of the following claims. First, upon her release, Mary acquires new beliefs about her colour experiences. Second, these beliefs concern facts she already knew before her release. The central tenet of this response is the idea that Mary acquires new concepts, usually called *phenomenal concepts*, about colour experiences that she could not possess before her release. These concepts enable her to have new thoughts and thus new beliefs about colour experiences.

My account of the two ways of thinking reply is distinctive in two respects. Firstly, I offer a motivation for this strategy. Many have advanced this response to block the knowledge argument. However, as some opponents and supporters have recently started to realise, this reply might be charged with being an *ad hoc* strategy.⁶ This means that the reply needs independent support. After outlining the central requirement that this strategy places on phenomenal concepts, I consider how this support can be provided.⁷ Preliminarily, I deny that this account can be based on the identification of phenomenal concepts with indexical concepts.

⁶ See, for this criticism Levine 2001, p. 86.

⁷ Recent attempts to justify the two ways of thinking strategy can be found in Perry 2001, Papineau 2002, Carruthers 2000, Tye 2003, Ayedede and Güzeldere 2003.

John Perry has recently provided a very articulated version of the two ways of thinking reply.⁸ The central idea in this view is that Mary acquires *indexical knowledge* about colour experiences that she could not have had before her release. Specifically, once she has colour experiences, she is able to think about them by means of the demonstrative concept [this]. Perry's account is based on an independently motivated theory of the individuation of the content of beliefs where indexical concepts figure. Moreover, it satisfies the central requirements that the two ways of thinking strategy puts on phenomenal concepts. Nevertheless, I argue that this account proves to be unsatisfactory.

The second distinctive element of my formulation of the "two ways of thinking" strategy derives from the view on introspection defended in this research. I argue that this strategy should be grounded on the existence of *recognitional concepts* of colours. In order to possess and apply these concepts there is no need to have other concepts or beliefs. Having colour experiences is sufficient. The novelty of Mary's phenomenal concepts concerning colour experiences "is parasitic" on the novelty of these colour concepts. Therefore, by acquiring recognitional colour concepts, Mary acquires new ways of thinking about the types of colour experiences she is having.

4 The Plan of the Volume

In Chapter 1, I offer the *modest reductionism hypothesis* as a plausible formulation of the idea that science can accommodate conscious experiences. Chapter 2 begins with Jackson's knowledge argument. After analysing this argument, I show that it is based on the unintelligible premise that Mary possesses a complete scientific knowledge. However, what emerges from this discussion is that that the general

⁸ Perry 2001.

strategy involved in this argument might still be viable to target the modest reductionism hypothesis.

Chapter 3 is concerned with the formulation of an intelligible version of the knowledge argument. By considering contemporary psychophysics and neuroscience, I characterise Mary's scientific knowledge. Specifically, this characterisation satisfies the general requirement of the modest reductionism hypothesis.

In Chapter 4, I focus on the knowledge that Mary supposedly acquires by having colour experiences. I first show that the standard account of the content of this knowledge needs justification. On this account, by seeing colours, Mary comes to believe that colour experiences have *qualia*. I show that the upholder of the knowledge argument should support this claim. However, I argue that this support cannot derive from the assumption that Mary is directly aware of her colour experiences and their features.

Chapter 5 considers a more plausible explanation of how Mary can form beliefs about the types of colour experiences that she has in virtue of seeing coloured objects. This claim is based on an inferential account of introspective knowledge of colour experiences. A characterisation of the beliefs that figure in Mary's supposedly new knowledge derives from this account. She has to discover something about the colour an object looks to have to her, in order to discover that a colour experience of a certain type has a certain property.

Chapter 6 considers, firstly, an important challenge to the idea that Mary might acquire new beliefs about colour experiences. The promoters of the *ability reply* have argued that Mary does not acquire any new belief when she sees coloured objects. They claim that knowing what it is like to have a colour experience is just possessing a certain ability to imagine, remember and recognise the experience. Specifically, I will show that the supporter of the knowledge argument can challenge this reply by endorsing a certain principle for the individuation of beliefs.

Therefore, we can concede to such a supporter that Mary acquires new propositional knowledge about her colour experiences.

In the final chapter, I consider whether Mary's acquiring new propositional knowledge implies the existence of facts that escape scientific knowledge. Many philosophers have answered to this question in the negative by offering what is usually called the *two ways of thinking reply*. I will endorse this reply and show how it can be based on the existence of purely recognitional concepts concerning colours.

1 A Hypothesis about Conscious Experiences

1.1 Introduction

Physicalism has offered an influential version of the hypothesis that conscious experiences can be accommodated in scientific terms. Physicalists maintain that physics should have a fundamental role in the study of mind. They assume that physics will unify the science that will provide an exhaustive account of the mind. In the present Chapter, I investigate how such an idea can have a precise and substantive formulation.

Section 1.2 considers a problem that appears to threaten the very possibility of formulating physicalism. Characterising the notion of being physical appears to be a central requirement for articulating this doctrine. Nevertheless, it has been argued that the physicalist who wants to use this characterisation faces a dilemma. I will illustrate this difficulty and how the physicalist can escape it.

Section 1.3 investigates another problem that physicalists have to consider when articulating their position. A programme for scientific unification lies at the core of physicalism. A traditional version of this programme is *classical reductive physicalism* as exemplified in philosophy of mind by the type-identity theory. Reductive physicalists claim that the ordinary understanding of the mind and/or its scientific successors will be reformulated in neuroscientific terms and ultimately by physics. They think that this hypothesis is authorised by an observable trend in the evolution of contemporary science. Specifically, reductionists claim that types of mental states are identical to types of physical states in the brain. Thus, all the sound psychological explanations, once couched in neuroscientific and physical terms, will be derivable within neuroscience and physics. However, I will illustrate

effective criticisms of reductive physicalism. These objections attack the notion of intertheoretic reduction implicit in this form of physicalism.

Section 1.4 will show that, although strong reductionism is untenable, physicalists can still promote a plausible programme for the unification of the study of the mind to physical science. Therefore, we can provide a plausible formulation of the hypothesis that science can give an account of conscious experiences.

1.2 Defining “Being Physical”

We need a precise formulation of the hypothesis that conscious experiences can be completely accounted for in scientific terms. Physicalism represents a very influential formulation of this claim. Physicalists adhere to a programme for the unification of the investigation of mind with the scientific study of the physical world. In their view, the mental can be completely described and explained in scientific terms. Specifically, psychology, biology and neuroscience can give such a scientific account. All these disciplines are regarded as potentially unifiable with physics. A wide debate concerns the issue of characterising this notion of unity.¹ However, before formulating such a unifying physicalist programme, one must first answer a more fundamental question.

This section considers how physicalists can characterise the ‘physical’. Many have argued that physicalists have to face a dilemma when attempting to answer this question.² According to this objection, articulating physicalism in terms of scientific knowledge renders it either false or unintelligible. I will show that this objection is resistible. Although certain strong formulations of physicalism are undermined by this criticism, a weaker version can withstand it. However, before

¹ See at p. 25.

² This dilemma was originally offered by C. G. Hempel, see Hempel 1980. A similar line of reasoning can be found in Crane and Mellor 1990 and in Montero 1999.

addressing this issue, we need a preliminary characterisation of the main assumptions involved in physicalism.

The physicalist thesis that mental states are part of the physical world is based on certain epistemological and ontological assumptions. These assumptions are expressed by a hypothesis that has shaped contemporary physicalism. Following David Lewis, we can call it the *explanatory adequacy of physics hypothesis*.³ This claim is: “The plausible hypothesis that there is some unified body of scientific theories, of the sort we now accept, which together provide a true and exhaustive account of all physical phenomena (i.e. all phenomena describable in physical terms).”⁴ Lewis expands on this as follows:

They are unified in the sense that they are cumulative: the theory governing any physical phenomenon is explained by theories governing phenomena out of which that phenomenon is composed and by the way it is composed out of them. The same is true of the latter phenomena, and so on down to fundamental particles or fields governed by a few simple laws, more or less as conceived of in present-day theoretical physics. (Lewis 1966: 23)

From the epistemological point of view, the main idea expressed in this passage is that all physical phenomena can be explained in terms of the laws regulating the behaviour of the fundamental particles posited by physics. This thesis is related to the ontological view that every phenomenon explained by a certain scientific theory, which differs from theoretical physics, is constituted by entities that ultimately are composed of the basic entities posited by physics.⁵

³ For a history of this principle, see Papineau 2002, pp. 232-256.

⁴ Lewis 1966, p. 23.

⁵ For the different ways in which physicalists would weaken this formulation of physicalism, see section 1.3, p. 25.

The phenomenon of digestion can be used to illustrate the explanatory adequacy of physics hypothesis. Digestive phenomena are constituted by chemical phenomena involving interactions between the molecules of certain substances. Thus, chemistry can explain digestion in terms of chemical laws that regulate the behaviour of these molecules. Now, the explanatory adequacy of physics hypothesis requires that the chemical laws governing the behaviour of these molecules can be explained in terms of laws concerning the atoms that constitute these molecules. In turn, there are basic explanations of the behaviour of atoms. Eventually, this descending chain of decomposition toward more basic components and explanations will end with the entities (and the laws that govern their behaviour) that are posited by theoretical physics.

The explanatory adequacy of physics hypothesis entails that if a physical phenomenon is explained by means of another phenomenon, then this latter phenomenon has to be physical.⁶ In particular, in Lewis's account, the hypothesis of the explanatory adequacy of physics appears to amount to the assumption that only physical phenomena can be causally efficacious with respect to other physical phenomena.⁷ The explanatory adequacy of physics does not entail the thesis that everything is physical. Therefore, it does not imply directly that the mental is physical; another premise is required.

Beside the explanatory adequacy of physics, physicalists suggest that the mental is causally efficacious with respect to the physical. Again, we can illustrate this premise by referring to Lewis. He argues that mental states cause physical effects and, thus, they figure in the explanation of physical facts.⁸ The explanatory

⁶ Lewis 1966, pp. 23-24.

⁷ Some physicalists directly assume the thesis of the causal closure of the physical avoiding any reference to explanatory adequacy. See, for instance, Papineau 2002, pp. 17-18.

⁸ See Lewis 1966, pp. 19-22. Lewis endorses and elaborates further the idea, suggested in Smart 1959, that an *a priori* analysis of *concepts* concerning experiences reveals that these states play certain *causal roles*. Other physicalists maintain that the mental is casually efficacious with respect to the physical as a matter of fact, see Papineau 2002, pp. 38-39.

adequacy of physics hypothesis states that only physical phenomena can explain other physical phenomena. Therefore, mental states are physical phenomena. For example, he maintains that pain is a mental state that causes certain behaviours of avoidance. Such behaviours consist of movements that can be described in terms of sciences such as physiology. For instance, the movement consequent upon painful interactions with the environment can be described in terms of the activation of certain motor neurons and certain modification in determinate muscles. Thus, pains play a role in the explanation of certain physical phenomena. Therefore, given the explanatory adequacy of physics, pain is a physical phenomenon. More specifically, Lewis thinks that pains and other sensations are physical states of the brain.⁹

So far, I have sketched two principal premises involved in an influential form of physicalism. First, physics is explanatorily adequate; meaning that only physical phenomena can explain other physical phenomena. Second, mental states figure in the explanation of certain physical phenomena. Now, I will investigate an implicit assumption involved in each of these premises.

The physicalist view delineated above involves what can be called a *theory-based conception* of being physical.¹⁰ Physical phenomena are defined by reference to physical science. According to the principle of the explanatory adequacy of physics, only phenomena composed of the fundamental particles posited by physics and explainable by means of the laws governing these particles are physical. However, this characterisation of being “physical” has been criticised.

Formulations of physicalism that involve the theory-based conception of being physical face a difficulty usually known as *Hempel's dilemma*.¹¹ This dilemma is taken to threaten the only two ways in which a theory-based conception of the physical can be formulated. Either the definition of “physical” is based on current

⁹ Lewis 1966, p. 24.

¹⁰ I take the name of this account from Stoljar 2001.

¹¹ Hempel 1980, a similar type of criticism can be found in Crane and Mellor 1990.

physical theory or it is grounded in some ideally complete future (or possible) physical theory. Geoffrey Hellman describes effectively the dilemma emerging from these two options:

... either physicalist principles are based on current physics, in which case there is every reason to think they are false; or else they are not, in which case it is, at best, difficult to interpret them, since they are based on a “physics” that does not exist - yet we lack any general criterion of “physical object, property, or law” framed independently of existing physical theories. (Hellman 1985: 609)

Let us examine the horns of this dilemma in the case of the explanatory adequacy of physics.

According to the first horn of the dilemma, if we characterise “physical” in terms of contemporary physics, then the thesis of the explanatory exhaustiveness of physics might turn out to be false. In this case, the principle requires that every physical phenomenon can be explained in terms of the ultimate particles and laws suggested by contemporary physics. This amounts to the assumption that contemporary physics provides a definitive inventory of physical reality. Nevertheless, as Barbara Montero puts it:

... if the physical is defined over current microphysics, and a new particle is discovered next week, the particle will not be physical. In addition, this is a consequence most philosophers want to avoid. (Montero 1999: 188)

The central idea is that the possibility that physics might require the introduction of new particles, or even that it might undergo radical theoretical revolutions, cannot be excluded *a priori*. In particular, the historical evidence seems to point to the contrary. The history of physics has seen radical theoretical revolutions and the inclusion of new entities. For example, the physics of the eighteen-century

mechanics had to be supplemented by the theory of electricity and magnetism. Moreover, in contemporary physics the nature of the ultimate components that exist at sub-atomic level is still an open question. Thus, the claim that every physical phenomenon can be explained by reference to the particles posited by contemporary physics might turn out to be false.

The other horn of the dilemma concerns formulating the theory-based conception of being physical by referring to a future (or possible) complete physics. In this case, two problems emerge. One difficulty is that the theory-based account cannot accomplish its main task. This conception does not give any precise content to the notion of “physical”. We cannot predict what entities or laws an ideally complete (future or possible) physics might refer to. The plausibility of the hypothesis of the explanatory adequacy of future physics is an empirical issue. Therefore, we cannot evaluate this claim before we possess such a scientific theory. At the present, we should be agnostic about the issue.

There is a second problem with the appeal to future complete physics. This difficulty becomes apparent when we consider the physicalist solution to the mind-body problem. The claim that mental entities can be completely explained in physical terms might just become a truism. We cannot exclude the possibility that a physical theory will involve reference to irreducible mental properties. Of course, stating this mere logical possibility is not a problem for physicalists. They might just claim that complete physics might not need reference to mental entities. However, some argue that completing the present theoretical physics might require reference to mental conscious states.

At the core of contemporary quantum mechanics lies a fundamental problem. The mathematical apparatus of quantum mechanics delivers accurate predictions in the realm of microphysics. The fundamental principle of quantum mechanics is the Schrödinger equation. This differential equation predicts the dynamics of wave functions that describe the basic particles. This principle requires that the properties

of basic particles, such as their position or momentum, may not always have well-defined values. However, for instance, when we measure the position of a particle, we find a definite value and not the combination of values required by the Schrödinger principle. For this reason, what is called the *measurement principle* is introduced. This principle states that, when we observe particles, the wave function does not behave as predicted by Schrödinger principle. Instead, it “collapses” in a way that the determinate value of properties such as the position or momentum of a particle can be established. The central problem of the interpretation of quantum mechanics is to explain why both these principles are required.

Many have argued that the solution to the problem of the interpretation of quantum mechanics requires referring to the conscious states of the observer. David Chalmers has recently advanced a proposal of this type.¹² He has argued that in reality the only principle required for quantum mechanics should be the Schrödinger equation. In addition, the effects described in terms of the measurement principle find a better explanation if we assume the existence of non-reducible conscious states. Of course, here I am not endorsing this solution to a very hard problem that lies at the core of contemporary physics. However, the fact that scientists and philosophers offer arguments for including consciousness in our present physical description of reality renders more pressing the problem that derives from formulating physicalism in terms of future developments of physics.

If Hempel's dilemma cannot be evaded, endorsing the theory-based characterisation of being physical implies that the problem posed by conscious experience is unsolvable; for we cannot even coherently formulate it.¹³ Noam Chomsky presents clearly this idea in the general case of the mind-body problem. He points out that Descartes could formulate this problem meaningfully because he

¹² Chalmers 1996, pp. 333-357.

¹³ See Montero 1999 and Levine 2001.

had a determined notion of body to oppose to that of mind. Notoriously, he assumed that the body was extended and subject to the laws of contact mechanics. Nevertheless, it seems that the changes that have affected physics have left us with a less definite notion of the body, or of the physical:

What is the concept of body that finally emerged? The answer is that there is no clear and definite concept of body. If the best theory of the material world that we can construct includes a variety of forces, particles that have no mass, and other entities that would have been offensive to the “scientific common sense” of Cartesians, then so be it: We conclude that these are properties of the physical world, the world of the body. The conclusions are tentative, as befits empirical hypotheses, but are not subject to criticism because they transcend some *a priori* conception of body. There is no longer any definitive conception of body. Rather, the material world is whatever we discover it to be, with whatever properties it must be assumed to have for the purposes of explanatory theory. (Chomsky 1988: 144)

Finding Hempel’s dilemma convincing some have offered alternative accounts to the theory-based conception of the physical.¹⁴

Some philosophers have suggested what can be called an *object-based conception* of the physical.¹⁵ They have maintained that an entity is physical when it is of the same type as some entity that is taken to be paradigmatically physical.¹⁶ The central idea in this account is that the paradigmatically physical entities can be introduced without any reference to contemporary, future or possible science.

¹⁴ For reactions to Hempel’s dilemma see Hellman 1985, Melnyk 1997, Montero 1999 and Levine 2001. See also the discussion in Poland 1994, Chapter 3, pp. 157 ff.

¹⁵ I take this name from Stoljar 2001.

¹⁶ See also Papineau 1993, p. 30.

Jackson illustrates how the object-based account characterises physical properties and relations:

The physicalists can give an ostensive definition of what they mean by physical properties and relations by pointing to some exemplars of non-sentient objects – tables, chairs, mountains, and the like – and then say that by physical properties they mean the kinds of properties and relation needed to give a complete account of things like them. (Jackson 1998a: 7)

Accordingly, physicalists can simply formulate the notion of physical property based on an ordinary understanding of the features of non-sentient ordinary objects. However, this account appears to be problematic.

As Jackson acknowledges, the possibility of *panpsychism* threatens the object-based account of “physical”.¹⁷ Panpsychists believe that every entity has a mind. On this view, the paradigmatically physical objects referred to in the object-based account of “physical” might have a mind. Nevertheless, the object-based account needs the existence of non-sentient objects. Therefore, the promoters of his view have to exclude the possibility of panpsychism. However, their theory does not seem to have any resource to specify further the nature of these non-sentient objects. It seems that they are just excluding the possibility of panpsychism by definition. Nevertheless, although panpsychism might strike us as completely implausible, its possibility cannot be ruled out just by stipulation. It seems that we need some substantive account of the nature of ordinary objects in order to exclude that they have a mind. This difficulty points to a more general one.

The difficulty created by the possibility of panpsychism is just a manifestation of a deeper problem in the object-based conception. This account of “physical”

¹⁷ Jackson 1998b, p. 7.

depends on the idea that we can have an ordinary understanding of what type of properties might figure in a complete account of objects such as chairs and tables. Nevertheless, our ordinary understanding of these properties might turn out to be inadequate. We have seen that the possibility of panpsychism entails that amongst the features of chairs, or tables, might be included mental properties. Even without contemplating such a possibility, the problem emerges. It is enough to consider the image of reality provided by contemporary physics. The ultimate particles, properties and laws that this science invokes in providing a complete account of ordinary objects are very different from those we can contemplate in our ordinary experience.¹⁸ Thus, a form of physicalism based on the object-based account can even contrast with physics. Therefore, endorsing such a theory requires countenancing the possibility that many of the basic assumptions of contemporary physics are false. However, not many physicalists can be willing to develop an account of physical reality alternative to that provided by contemporary physics.

The problem that afflicts the theory-based and object-based conceptions of being physical might be avoided by denying an assumption that they share. This is the idea that a formulation of physicalism needs an adequate characterisation of physical entities.¹⁹ Those who follow this strategy maintain that we have a better understanding of the mental than the physical. Joseph Levine, for example, claims that our mental life involves *phenomenal* and *representational* properties. Here it is sufficient to say that according to him, phenomenal properties characterise conscious mental states. Moreover, he assumes that mental states that represent the world as being in a certain way have representational features. Thus, he suggests that physicalism (in his words “materialism”) is the theory that gives ontological priority to non-mental properties:

¹⁸ A criticism of this type can be found in Levine 2001, p. 20.

¹⁹ This account is suggested in Montero 1999. See also Levine 2001, pp. 20-21.

M': Only non-mental properties are instantiated in a basic way; all mental properties are instantiated by being realized by the instantiation of other non-mental properties. (Levine 2001: 21)

Even without explaining the notion of realisation involved here, it emerges that this statement is a formulation of physicalism that does not involve reference to any physical theory. Levine clarifies this formulation of physicalism as follows:

It is not important for purposes of this thesis whether we have an adequate conception of what these basic non-mental properties are, so long we're clear that they are not representational or phenomenal. If a future physics tell us that among the basic properties of elementary particles or fields are representing quantity x or feeling pain, then materialism is false. (Levine 2001: 20)

Defining the physical via this negative approach clearly avoids the problems of the theory-based and object-based accounts. However, is this strategy satisfactory?

The negative definition of being physical appears to be inadequate. Although this characterisation avoids Hempel's dilemma, it renders physicalism completely independent of the practices and achievements of current sciences. In this case, the physicalist solution to the mind-body problem amounts to the claim that non-mental properties have an ontological and explanatory priority over mental properties. Therefore, physicalism does not imply the endorsement of any specific scientific research project. This means that the philosophical discussion of the mind-body problem should be independent of any current scientific attempt to explain and describe mental properties.²⁰ Nevertheless, this appears to contrast with the general attitude of many physicalists within philosophy of mind. Specifically, such a way of

²⁰ For example, Thomas Nagel characterises scientific knowledge in terms of a notion of objectivity understood as independence from subjective points of view Nagel 1974 and Nagel 1986.

defining the physical undermines one of the main reasons offered by physicalists for their position.

Physicalists place confidence in the explanatory power of scientific knowledge, given the actual developments of contemporary science. In particular, many physicalists have been impressed by the results of biology and neuroscience that have explained many aspects of both normal and pathological human behaviour.²¹ Thus, many physicalists want to offer a conception of the mind that is not only consistent with contemporary science, but that might aid scientific progress too.²² Clearly, justifying physicalism by referring to the current scientific practice must involve a theory-based conception of physical. However, this leads us back to Hempel's dilemma. Therefore, we have to see whether the physicalist can meet this difficulty without abandoning the theory-based conception of being physical.

One of the horns of Hempel's dilemma states that if "physical" is defined by reference to future physical science, we lack any substantive grasp of the explanatory adequacy of physics. This objection might be resisted by providing a philosophical account of scientific knowledge at a level of generality that can characterise it independently of its future changes.²³ However, a less demanding enterprise is to see whether the other horn of Hempel's dilemma can be met.

If physicalism involves the thesis that *all* physical phenomena are explainable in terms of the entities and laws posited by current physics, then physicalism is false.²⁴ For we cannot exclude that future physics might consider new fundamental

²¹ David Papineau illustrates how the idea of completeness of physics is not a methodological or metaphysical principle based on *a priori* considerations. He argues that advancements in the understanding of neurophysiology due to biochemistry in the first half of twentieth century are central in establishing this principle. See Papineau 2002, pp. 232-256.

²² Fodor 1974, Smart 1959, Churchland 1986.

²³ One of these attempts is offered in Poland 1994.

²⁴ Another interesting defence of a theory-based account that refers to current physics is given in Melnyk 1997. Melnyk argues that although defining physicalism in terms of contemporary physics might render it false, endorsing this doctrine is still rational.

particles governed by laws that we do not know presently. However, physicalism in the philosophy of mind might be detached from the general statement that contemporary physics provides the ultimate catalogue of physical entities. If so, then it might follow that the second horn of Hempel's dilemma is not effective. Let us consider this option.

Physicalism in philosophy of mind can be formulated as involving the thesis that mental phenomena can be explained in terms of properties of the kind recognised by current physical science. Frank Jackson has suggested this way of understanding physicalism. Dealing with the issue of defining physical properties and relations, he states that:

... they will be broadly of a kind with those that appear in current physical science, or at least they will be as far as the explanations of macroscopic phenomena go, and the mind is a macroscopic phenomenon. (Jackson 1998a: 7)

Let us consider a version of this view.


J. C. C. Smart suggests tying physicalism, understood as a theory about the mind, to contemporary physics.²⁵ A central assumption in this proposal is that current physics of "ordinary matter" is complete. This means that a class of macroscopic phenomena can be completely described and explained in terms of principles and properties of current physics. Smart illustrates this assumption by referring to the position of the physicist Gerald Feinberg. This scientist is reported to claim:²⁶

That the theory of the electron, proton, neutron, neutrino and photon and their anti-particles, when they have such, is enough to explain the properties of ordinary matter (Not what goes on inside neutron stars or

²⁵ Smart 1978, Smart 1989.

²⁶ Feinberg 1966.

inside black holes, or the behaviour of the transitory particles created only with big cyclotrons). Feinberg thus holds that the “Thales Problem” (of what the world of familiar objects is made of) has essentially been solved. (Smart 1989: 81)

Smart concedes that there will be changes in physics. However, he claims that these changes will affect only the physics of certain phenomena. We can expect changes in the theories concerning phenomena at sub-atomic level that are studied under special laboratory conditions. Moreover, we can expect transformation in those theories that consider the universe  the large. Nevertheless, similar changes will not affect the scientific descriptions and explanations of macroscopic phenomena that involve ordinary matter. For example, there will not be discoveries that alter the fact that the hydrogen atom contains one proton and one electron and that water is H₂O.

The second tenet in Smart's position is that contemporary physics of ordinary matter can account for the mind. He assumes that: “The properties of mind depend on the properties of 'ordinary matter'”.²⁷ Specifically, he claims that the properties of the mind depend on those of the brain. Moreover, he states that:

The properties of the brain are those of assemblages of neurons, and the study of neurons requires only quite well known physics and chemistry.
(Smart 1989: 80)

Smart concedes that there are gaps in our understanding of the functioning of the brain and of the mind. However, they derive from difficulties we are facing in the detailed application of known principles of chemistry and physics.²⁸ Therefore, we

²⁷ Smart 1978, p. 341.

²⁸ Smart 1978, p. 349.

should not expect that “discoveries about quarks, black holes, theories of strings and superstrings” might affect our knowledge of the mind.²⁹

Smart suggests that the idea that scientific knowledge can accommodate the mind is a substantive doctrine that is not obviously false. This is achieved by formulating physicalism properly. Such a formulation requires detaching physicalism from the thesis that contemporary physics provides an account of all physical entities. A more plausible position is that current physics is explanatorily adequate for a class of macroscopic physical phenomena. This thesis is then coupled with the idea that the mind is one of these macroscopic phenomena. Thus, we have a plausible formulation of the main intuitions involved in the physicalist solution to the mind-body problem. However, this formulation might need some refinement in order to be completely satisfactory.

One initial difficulty stems from the fact that many physicalists would deny that mental properties depend only on those of the brain. This is because they support an externalist account of the conditions that individuate mental states.³⁰ On this view, certain relations of an individual to her environment figure in the conditions that specify her mental states. However, it seems that many of these authors appear to acknowledge that physics might be able to describe and explain these causal relations. Similarly, they appear to extend this hypothesis to the relevant features of external objects on which mental properties might depend. However, specifying this relation of dependence might create another difficulty.

Philosophers such as Smart and Lewis have formulated physicalism as a reductionist doctrine. Reductionism involves interrelated ontological and epistemological theses. In philosophy of mind, reductionists state that types of mental phenomena are identical to types of physical phenomena. On the

²⁹ Smart 1989, p. 80.

³⁰ In particular, externalist accounts of conscious experiences have been proposed in Dretske 1995 and Tye 1995.

epistemological side, reductionists account for reduction as a relation between scientific theories. Consequently, they claim that all the scientific theories reduce, in this sense, to physics. However, as we will see in the next section, there are reasons to reject this formulation of physicalism.

1.3 The Limitations of Classical Reductionism

In our quest for a precise formulation of the hypothesis that conscious experiences can be an object of scientific knowledge, we have investigated physicalism. It has emerged that one plausible version of physicalism is based on the thesis that mental states can be completely described and explained by a science that can be unified by the study of macro-phenomena, as investigated by contemporary physics. We have seen that physicalists in philosophy of mind do not need to endorse a general statement about the nature of all physical phenomena. They can just assert that mental phenomena can be exhaustively understood in term of the physics of macro-phenomena involved in the biological processes of the brain. This weaker form of physicalism seems to constitute a substantive position that does not appear to be obviously false. However, it has to

be seen how the central notion of unification is to be understood. Clearly, without such an understanding we cannot substantiate the claim that conscious experiences can be accommodated by scientific knowledge.

In this section, I consider whether there is a plausible formulation of the project for unification involved in physicalism. Physicalists have articulated different versions of this project by endorsing different epistemological and ontological theses. From the epistemological point of view, physicalists differ in their accounts of the relation that should exist between the scientific theories studying the mind and the physical world. On the ontological side, they have promoted different views about the relation between mental and physical entities consistent with their epistemological assumption.

Firstly, I will illustrate reductive physicalism; a theory that can be regarded as a classical formulation of contemporary physicalism. The upholders of this version of physicalism formulate their project for unification by maintaining that scientific theories of the mind are reducible to physics. A certain model of inter-theoretic reduction has guided the formulation of this doctrine. Given this model, the project for unification amounts to the hypothesis that there are *bridge laws* that connect the vocabularies of the different sciences in a descending order with the language of physics as its first element. Usually, the defenders of this approach have maintained the ontological thesis that mental entities are identical to physical ones.

Secondly, I will consider some influential attacks on this way of formulating the physicalist programme for scientific unification. I will outline the objections of those who have maintained that there cannot be bridge laws connecting psychological and neuroscientific vocabularies. Lacking such laws, the project of reduction cannot be carried forward. However, although these criticisms show certain limitations in the reductionist programme, they should not be taken as conclusive proof that the physicalist programme (for the unification of the study of the mind by physics) is misguided.

An influential version of contemporary physicalism has expressed the physicalist programme in the context of reductionism. According to reductionists, there is a trend in contemporary science that justifies the idea that ideally all scientific knowledge is reducible to physical knowledge.³¹ Obviously, the notion of intertheoretic reduction is central in this position.

³¹ For a classical statement of this project, see Oppenheim and Putnam 1958.

An influential view on intertheoretic reduction focuses on the explanatory capacities of scientific theories.³² According to this account, a theory T_2 is reduced to a theory T_1 when, amongst other conditions, the data explainable by T_2 are explainable by T_1 .³³ Ernest Nagel has provided a classical account of such an explanatory subsumption of theories.³⁴ In his formal analysis, a theory T_2 is reduced to a theory T_1 when all the statements of T_2 are deduced from the statements of T_1 . When there are predicates and terms of T_2 that do not appear in T_1 , Nagel's model requires extra premises that connect the vocabularies of these theories. These additional assumptions, usually called *bridge principles*, are bi-conditional statements relating the statements of the reduced theory to certain statements of the reducing theory. For instance, let us consider two theories T_2 and T_1 that differ in their vocabularies. Theory T_2 is reducible to T_1 , when the statements T_2 are deducible from a theory obtained by adding bridge principles to T_1 . In Nagel's account, the reducing theory explains the reduced theory because of a *covering law model of explanation*. According to this model, an *explanation* is an *argument* (deductive-nomological or inductive-statistical) whose *premises* include laws and *conditions*, and whose conclusion is a description of the phenomenon to be explained. According to Nagel's model, all the laws of the reduced theory are logical consequences in the reducing theory; therefore, the explanatory power of the first theory is transferred to the latter.

³² It is important to signal a different attempt to characterise the idea of the unification of scientific knowledge by means of reductionism. This is a programme for semantic reduction. On this view, all the scientific sentences could be proved logically equivalent to observational sentences via the definition of scientific terms by means of observational terms. For this reductive programme, see Trout 1991, pp. 388-389.

³³ See Oppenheim and Putnam 1958.

³⁴ Nagel 1961, pp. 337-397.

Type identity theory exemplifies reductive physicalism in philosophy of mind.³⁵ The central thesis of this position is that types of mental states are identical to type of brain states. This ontological assumption implies that statements referring to certain mental facts are equivalent to certain statements expressing physical facts about the brain. For example, if the type mental entity pain is identical to the neural type *C-fibres* firing, we have the bridge principle stating that an individual is in pain if and only if she has her *C-fibres* firing.

Type identity theory amounts to the idea that science of the mind, either as ordinary psychology or some scientific elaboration of it, will be completely formulated in neuroscientific terms. From the ontological point of view, this means that neuroscience refers to the mental entities posited by psychology. Similarly, at the epistemological level, the explanations couched in psychological terms can be reformulated as neuroscientific explanations. The strong reductive picture offered by type identity theory has been criticised even by philosophers who endorse a form of physicalism. Let us consider a criticism advanced by the supporters of functionalism.³⁶

Functionalists have argued that there cannot be bridge laws connecting mental and physical theories. Therefore, type identity theory is wrong. They maintain that each mental state is defined by a certain causal role. For instance, a functionalist analysis of pain might state that being in pain is being in a state that is caused by certain stimuli and that causes certain responses of avoidance and certain other mental states. Now, according to functionalism, different physical states can play the same causal role. Even physical systems as diverse as human brains, animal

³⁵ This theory was promoted in Smart 1959, Place 1956, Feigl 1967. See also Feigl 1934 for a precursor of this doctrine. These authors did not mention explicitly any account of intertheoretic reduction. However, their position appears to imply such a model. Moreover, Fodor and Putnam mounted influential objections to type identity theory understood as involving this account of reduction; see Fodor 1974 and Putnam 1967.

³⁶ Donald Davidson, who endorses a form of physicalism, has provided other influential arguments against the type-identity theory. See Davidson 1970.

brains, and computers can realise the same causal role that defines a certain mental state. Given this *multiple realisability* of mental states in different physical states, there cannot be bridge laws connecting psychology and neuroscience. Therefore, the possibility of type identity between mental and physical entities is ruled out and reductionism is unattainable.³⁷

The main upshot of this type of criticism is that psychology is autonomous from physics. In fact, psychological laws and explanations concern certain kind of entities that cannot be type identical with physical kinds. It is important to add that despite these antireductionist theses, most functionalists endorse a weak form of physicalism as the doctrine that tokens of mental states are identical to tokens of physical states. However, it can be legitimately asked whether this antireductionist conclusion affects the physicalist programme for unification. First, let us consider how many reductionists responded to these objections.

Many reductionists in philosophy of mind have reacted to the objection of functionalists. Some of these replies centre on the idea that causal analyses of mental states are compatible with the existence of bridge laws connecting sentences belonging to psychology to those belonging to neuroscience. One strategy is to argue that although in general there are not bridge principles that connect mental states, individuated by way of causal roles, to type physical states, there are, nevertheless, local reductions when we consider the psychology of certain species. According to these philosophers, although a mental state such as pain can have different realisations in different physical systems, when we consider the human species it can be maintained that there is only one type of brain state that realises

³⁷ This type of argument can be found in Fodor 1974, Putnam 1967. More precisely, the core of this objection is the idea that the conjunctions of heterogeneous physical realisers that play the causal role that define mental states cannot figure in physical laws. See on this Fodor 1974, p. 123 and Kim 1992, p. 318.

that mental state. Thus, for a human being having that brain state is a necessary and sufficient condition for having pain.³⁸

Antireductionists, in turn, have reacted by arguing that the mental states are multiply realisable within the same species, and even in the same individual at different times.³⁹ Counter-replies have stressed that these arguments, if they show anything, point to the fact that reductions should be localised at the level of certain individuals or even at the level of individuals at certain times. Nevertheless, the possibility of bridge principles connecting types of mental states and types of brain states remains open and individuating these principles is a worthwhile scientific enterprise.⁴⁰ Thus, it seems that reductionists in philosophy of mind might even be able to formulate their hypothesis against the objection of functionalists. However, by just focussing on this debate we might miss a more general problem that might afflict reductionism.

A deeper problem for the kind of reductionism considered here derives from its connection with the Nagelian model of reduction. Many have advanced a reductive programme as a plausible hypothesis based on a trend empirically ascertainable in the development of science. Thus, it makes sense to ask whether this notion of intertheoretic reduction is supported by cases in the history of science. However, the development of philosophy of science has shown that Nagel's account is less than satisfactory. In particular, there are two orders of reasons that can be offered to support this criticism.⁴¹ First, it can be argued that, amongst all the cases of historical scientific reduction, very few appear to approximate Nagel's model. Second, even when a Nagelian reduction might be available, there is no requirement

³⁸ This strategy is pursued in Lewis 1972, Kim 1992, Enç 1983.

³⁹ Tye 1983 and Hornsby 1984.

⁴⁰ Jackson, Pargetter, and Prior 1982. See, for a reply, Hornsby 1984.

⁴¹ These two lines of argument are pursued in Smith 1992, pp. 28-32.

for an ontological and explanatory absorption. Let us illustrate the first type of considerations.

In the 1960s, when Nagel's account of reduction was assuming a central position in philosophy of mind, many philosophers of science started to question it.⁴² These authors highlighted the fact that within the history of science, although many episodes can be regarded as successful cases of reduction, none in fact fit Nagel's proposal. Some noticed that in many cases the reducing and the reduced theory were not logically consistent.⁴³ For example, Galileo assumed falsely, that falling bodies have uniform vertical acceleration, whatever finite interval we consider. If Newtonian mechanics had derived this principle, according to Nagel model, it would have been false.⁴⁴ Even the reduction of classical thermodynamics to statistical mechanics, used by Nagel as a paradigmatic illustration of his model, was shown not to fit his account of reduction.⁴⁵ Moreover, others argued forcefully that the succession of scientific theories consisted instead of a total or partial replacement of the old theory's ontology with a new one.⁴⁶ Bridge laws do not come into this picture. The relation between modern chemistry and the phlogiston

⁴² The adequacy of this model in dealing with historical cases of successful scientific reductions has been questioned. Some criticisms came from Feyerabend 1962 and Hooker 1981. See, for a clear illustration of these objections and alternative models of intertheoretic reduction, Bickle 2003, and Bickle 1998. In addition, it is important to notice that Nagel's theory of reduction is based on the covering law model of explanation. Thus, indirect criticism of his theory might come from the problems inherent in his account of explanation. See, for the debate generated by this account of explanation, Salmon 1990.

⁴³ Schaffner 1967.

⁴⁴ See Schaffner 1967.

⁴⁵ Hooker claims that the laws of equilibrium in thermodynamics can be derived from statistical mechanics only when certain limit conditions are realised. Moreover, those limits can never actually be realised. Therefore, he concludes: "In a fairly strong sense thermodynamics is simply conceptually and empirically wrong and must be replaced" (Hooker 1981: 4).

⁴⁶ In Feyerabend 1962 it is maintained that successive scientific theories are ontologically incommensurable and thus there is no way to find an illuminating formal relation between them. However, many considered his position extreme and tried to elaborate the model in the spirit of Nagel's analysis that would illustrate some formal relation between successive theories. See Bickle 1998, p. 27.

theory illustrates a case of complete ontological replacement. In this case, there is match between de-phlogisticated gas and oxygen. Thus, we can explain why the phlogiston theory worked as well as it did. However, there cannot be bridge laws that connect the ontologies of the two theories. Another example is given by relativistic mass. This mass is conceived as a relation of an object with countless frames of reference. However, in the case of the classical conception, mass is a monadic property of an object.⁴⁷

The upshot of these debates was the realisation that an account of reduction should accommodate a wide spectrum of cases. Paul Churchland has called *bumpy* and *smooth reduction* the two extremes of this spectrum.⁴⁸ Bumpy reductions occur when many elements of the theory considered for reduction cannot be accommodated within the reducing one. In certain cases, bumpy reductions amount to complete replacements. This means that it is not possible to maintain the principles or the explanatory structure of the old theory. When a theory reduces smoothly to another one, it is possible to find some relevant correlation between their respective laws. When this mapping of the laws is perfect, the ontological result is the identification of the entities and the properties considered by the old theory with those of the new one. For example, in the case of physical optics and electromagnetic theory we have the identification of rays of light with electromagnetic radiation. As pointed out by Peter Smith, between these two extremes lie those cases where:

The correspondence between old and new is neither so tight as to sustain unqualified cross-theoretic identification, nor so loose as to make straight elimination of the ontology a comfortable option. (Smith 1992: 29)

⁴⁷ Another important type of criticism of Nagel's account of reduction is based on the idea that, in many cases, the reducing and reduced theories co-evolve, see Churchland 1986.

⁴⁸ Churchland 1985, p. 11.

Having seen the general requirement for an account of reduction, let us consider a model that appears to provide important insight into the nature of this relation between theories.

The claim that bridge laws should connect the vocabularies and the ontologies of the reduced and the reducing theory is too strong.⁴⁹ In fact, the resulting kind of reductionism appears to accommodate few episodes in the history of science. Instead, a model advanced by Clifford Hooker has been quite influential.⁵⁰ This model of reduction is committed to Nagel's idea that reduction involves deduction, but without requiring the existence of bridge laws. In this model of reduction, in the reducing theory T_b we have to construct an analogue T_r^* of the reduced theory T_r , adding to it a set of principles C_r . The set C_r contains various boundary conditions and some limiting assumptions, some of them counter to fact, that allow us to infer from T_b the theory T_r^* . Let us consider, for instance, the reduction of Galilean to Newtonian mechanics. Such a reduction requires adding to Newtonian mechanics a counterfactual condition. Namely: falling bodies' vertical acceleration is constant whatever their distance from Earth. In this case, there is a relation of deduction between T_b and T_r^* . However, this model differs from the previous theories of reduction. The difference is in the relation assumed between T_r and its representation T_r^* in T_b . On this account, T_r^* is not construed from the vocabulary of T_r . In fact, the relation between T_r^* and T_r is not given in terms of bridge principles. Instead, the existence of an *analogue relation* between these two theories is required. Reducing theories preserve an equipotent image of reduced theories

⁴⁹ For an account of the principal philosophical attempts at characterising intertheoretic reduction, see Bickle 2003.

⁵⁰ See Hooker 1981 and Smith 1992.

without a comprehensive mapping. Thus, what is reduced is a structure already within the vocabulary of the reducing theory T_b .⁵¹

This model of reduction appears to account for the possible cases in the *reductive spectrum*, considered above. On this account, the cases of reduction in this spectrum are characterised by different degree of the correspondence AR between the image T_r^* , embedded in the reducing theory and the reduced theory T_r . In the case of very *bumpy intertheoretic* reductions, this correspondence is minimal. On the other hand, in the case of very smooth reductions, the relation AR is so tight that we can assume that there is cross-theoretic identification of the entities postulated by the reduced theory with those postulated by the reducing one. Thus, it seems that there are reasons to accept a more general notion of reduction than the one suggested by Nagel. Moreover, philosophers have provided semiformal analyses of such a notion of reduction. However, we have seen that at one end of the reductive spectrum there are cases of reduction where the AR relation is very stringent. Then, we might assume that in these cases Nagel's requirements are satisfied. Let us now consider whether this is the case.

The cases of smooth reduction do not imply the absorption of one theory into the other. Peter Smith considers the case of elementary fluid mechanics.⁵² The basic laws of this theory can be obtained by the conjunction of those of classical

⁵¹ Paul Churchland, in Churchland 1985, illustrates in the following way this kind of reduction. Let us assume that T_r^* is a set of theorems of the restricted theory T_b , i. e. T_b plus the set of conditions C_r . Let us suppose that T_r^* contains these theorems:

(1) $\forall x A(x) \rightarrow B(x)$, (2) $\forall x (B(x) \wedge C(x)) \rightarrow D(x)$.

In addition, T_r contains these theorems:

(3) $\forall x J(x) \rightarrow K(x)$, (4) $\forall x (K(x) \wedge L(x)) \rightarrow M(x)$.

The theorems in T_r appear, in their syntactical structure, relevantly isomorphic to those in T_r^* . Thus the relation of analogy AR is based on similarities expressed in terms of the syntactical complexity of the reduced theory and its image in the reducing theory. It is important to notice that this is not meant to be a precise definition of the notion of analogue relation but just a way to gesture to the idea involved in Hooker's account. For a recent attempt at providing a precise characterisation of the notion of analogue relation, see Bickle 1998.

⁵² Smith 1992.

Newtonian mechanics, the laws of thermodynamics, and the assumption that fluids are substances that deform continuously under application of shear stress (even for a very small amount). By using a molecular theory of the fluid state, we might prove this assumption. Thus, in a manner very close to a Nagelian reduction, we might deduce the basic laws of fluid mechanics within a theory that combines Newton's laws, plus thermodynamics and the molecular theory of matter. However, this does not imply fluid mechanics is completely absorbed by the latter theories.

Smith argues that recent developments in the study of complex fluids system show that elementary fluid mechanics retains its explanatory autonomy from the molecular theory of matter. Thanks to the discovery of Lorenz equations, it is possible to study the behaviour of certain chaotic systems in a way that:

... is untouched by remarking that the underlying basic principles of fluid mechanics are themselves derivable from more fundamental theories. (Smith 1992: 31)

The principles of fluid mechanics can be derived from theories that study phenomena that are more basic. However, this discipline studies systems whose complexity has required *sui generis* modelling techniques and explanations. In particular, given the autonomy of this discipline; it might even be the case that the assumption that fluids are substances that deform continuously under application of no shear stress has different molecular explanation in different substances. Thus, it could be the case that the explanation for why this principle holds is different in the case of water than the case of oil. In other words, the actual scientific practice in the study of fluids could be consistent with the multiple realisability of the phenomena it investigates.⁵³

⁵³ Smith 1992.

The scientific practice suggests that the relation that might obtain between psychology, neuroscience and ultimately physics is less determinate than the one proposed by reductive physicalists. In reality, there is a spectrum of options. At the one end of this spectrum, there are cases where there is a complete ontological replacement. At the other end, cases appear to satisfy Nagel's requirements. However, even these latter cases do not support the strong conclusion that the reduced theory is superseded. Certain disciplines retain their explanatory autonomy.

To recapitulate, philosophers of mind and of science object convincingly to the idea neuroscience will reduce psychology by absorbing it. In the philosophy of mind, functionalists have argued that psychology is autonomous from neuroscience, and other physical sciences, given the very nature of mental states. Many philosophers of science have argued that Nagel's analysis does not capture the dynamic of actual episodes of successful reduction in science. Thus, it appears that physicalists who want to formulate their position in relation to scientific practice should avoid certain commitments. Namely, they cannot support their reductive programme in terms of the strong requirement of the ontological and explanatory absorption required by Nagel's account of reduction. Shall we, then, conclude that the physicalist has to renounce any idea of the explanatory adequacy of physics? Moreover, should she abandon the programme for the unification of the study of the mind with the rest of the study of the physical world?

The next section aims to show that a plausible form of physicalism can be formulated in terms of a weak reductionist programme that retains a plausible formulation of the aspiration to give centrality to physics and neuroscience. At the same time, such a form of physicalism does not require the strong and untenable commitments of the classical account of reduction provided by Nagel.

1.4 Modest Reductionism

In the previous sections, I argued that a plausible formulation of physicalism should satisfy two requirements. First, physicalism should provide an account of the notion of being physical in the context of philosophy of mind. This can be done if it is assumed that the study of the mental should be unified to the account of the “bulk matter” offered by contemporary physics. In particular, physicalists should not connect this thesis to a far stronger claim, namely, that contemporary physics provides the account of the ultimate physical reality at the subatomic level. Second, physicalists should avoid the strong commitments of reductive physicalism.

The previous two requirements appear to generate an inconsistency. How can physicalists maintain a project for the unification of the study of the mind to physics without endorsing a form of reductionism? The scope of this section is to illustrate a form of physicalism that might keep these two assumptions together. My main claim is that there is a way of formulating physicalism that is substantive and that preserves the central intuition of the prominent role that physics should have in the study of the mind.

We have seen that the hypothesis that psychology will reduce to neuroscience and eventually to physics, because mental entities are type identical to physical states, is too demanding. However, the general requirement for having intertheoretic reduction can be less stringent than having an ontological relation that supports bridge laws. As pointed out by Nagel himself the general idea is that inter-theoretical reduction:

... is the explanation of a theory or a set of experimental laws established in one area of inquiry, by a theory usually though not invariably formulated for some other domain. (Nagel 1961: 338)

This passage offers a general characterisation of inter-theoretic reductions that might not require to be spelled out in terms of bridge laws. Moreover, such an

account might not require the ontological identification of types of the entities of the reducing doctrine with those of the reducing one. The only requirement is that the reducing theory is able to explain why the reduced one works as well as it does. Let us see whether we can spell out this suggestion in more detail.

Peter Smith has articulated the general requirement for intertheoretic reduction in terms of the notion of *explanatory interfacing*. According to this proposal, a certain theory can provide explanations in terms of certain regularities discerned by using a taxonomy that is autonomous from that of sciences that are more fundamental. He assumes that explanation has a contrastive character; thus the fundamental form of an explanation is: “*that p (rather than $q_1, q_2...$) explains why r (rather than $s_1, s_2...$)*”.⁵⁴ Thus, it can be that the phenomenon r that has to be explained and the phenomenon p that explains it and the alternatives $q_1, q_2, \dots, s_1, s_2, \dots$ are described and individuated in ways that are not reducible to those of some other science.

Smith illustrates the notion of explanatory interfacing with the example of the explanation why a certain person wrote a certain cheque in a certain situation. Now, psychology can explain this by maintaining that the person’s desire to pay her bill and her belief that using cheques is a way of paying (and not her desire to see her bank account decreasing) causes her writing the cheque (and not using cash). This explanation requires certain patterns of events that are discerned by using psychological notions such as desires and beliefs. Now neuroscience cannot pretend to explain the patterns that psychology discerns because there it might no way to reduce the theory of beliefs and desires to neuroscience. However, Smith argues that this does not mean that neuroscience has no explanation to provide in this case. We already know that neuroscience can explain why the person wrote the cheque.

⁵⁴ This assumption does not appear to play any important role in his argument. The argument applies to any account of explanation that involves the idea that an explanation in a certain theory requires that the theory provides a description of its phenomena.

Physiological mechanisms explain the movements of the arm and the fingers that are required in writing the cheque. Thus, neuroscience is able to provide a description and explanation of the occurrence of an event that psychology re-describes and re-explains at a different level. This illustrates the kind of reductive relation that can be endorsed in a physicalist programme.

To sum up, the project for the unification underlying physicalism should not be articulated in terms of strong reductionism. Physicalists might endorse the more modest assumption that we might expect and try to provide interfacing explanation between psychology and neuroscience, which in turn might lead to explanatory interfacings of the study of the mind to contemporary physics.

1.5 Conclusion

Many philosophers think that phenomenal consciousness represents a problem for anyone attempting to account for the mental scientifically. However, before addressing the question whether there is such a problem we have to formulate the issue clearly. This requires a preliminary understanding of the hypothesis of those supporting the idea that science can provide a complete account of mental life.

I have investigated how we can characterise such a hypothesis in a plausible way. I have suggested that modest reductionism might provide a plausible formulation of this assumption. This is the idea that a scientific understanding of the mind can be provided by means of a science that can be explanatorily interfaced with the current physical study of macro-phenomena as explained by chemistry, biology and, ultimately, physics.

Given this characterisation of the idea that conscious experiences might be completely knowable in scientific terms, we have to consider whether we can provide a substantive account of the problem that phenomenal consciousness can present to such a thesis. In the next chapter, I will consider an extremely influential

argument that, if sound, would show that even the hypothesis of modest reductionism might be threatened by conscious experiences.

2 The Knowledge Argument

2.1 Introduction

The previous chapter introduced a central issue within contemporary philosophy of mind, namely, the problem of whether or not we should accept the hypothesis that science can account for conscious experiences. I suggested a plausible formulation of this statement, which I called the hypothesis of *modest reductionism*. This is the claim that conscious experiences can be completely accounted for in terms of a science that can be explanatorily interfaced with contemporary physics of macroscopic phenomena. In order to proceed, then, we must next determine what problems, if any, might afflict the hypothesis of modest reductionism.

This chapter illustrates the knowledge argument offered by Frank Jackson.¹ This is one of the most influential contemporary criticisms of the idea that conscious experiences can be accounted for in scientific terms.² My aim is to investigate whether Jackson's argument raises a difficulty for the hypothesis of modest reductionism. The main conclusion of the chapter is that, in order to carry forward this investigation, we should accomplish two tasks. First, we should provide a suitably revised version of the argument suggested by Jackson. Second, we should clarify further the hypothesis of modest reductionism.

Section 2.2 introduces the most influential presentation of the knowledge argument. This account involves Mary, a vision scientist who has a complete scientific knowledge of colour and colour vision but who has never seen colours. According to Jackson, if she experiences a colour she will learn *what it is like* to

¹ Jackson 1982 and Jackson 1986.

² The bibliography Chalmers 1999 lists fifty papers dedicated to this argument. Just to mention the most recent philosophical books on consciousness, extensive discussions of the knowledge argument can be found in Carruthers 2000, Papineau 2002, Perry 2001, Levine 2001, and Tye 2000.

have a colour experience and thus her scientific knowledge is incomplete. Before establishing whether the knowledge argument can determine a problem for the hypothesis of modest reductionism, we have to consider a criticism of this line of reasoning.

In section 2.3, I illustrate and endorse this objection as voiced by Daniel Dennett and Patricia Churchland.³ They have argued convincingly that we cannot grasp the complete scientific knowledge that Jackson ascribes to Mary.

In section 2.4, despite this criticism, I will maintain that the general strategy used in the knowledge argument might still be useful to test the hypothesis of modest reductionism. This requires reformulating a version of the knowledge argument that can target modest reductionism and resists the objection raised by Dennett and Churchland. However, I will argue that providing such a reformulation requires a more accurate characterisation of the hypothesis of modest reductionism.

2.2 A Thought Experiment about Mary

With a formulation of the hypothesis that science can accommodate conscious experiences in place, we can now start investigating whether such an assumption is problematic. Contemporary philosophers of mind have offered different arguments against the idea that consciousness can be studied in scientific terms. These arguments target different ways to spell out such a supposition.⁴ Amongst these objections, Frank Jackson's knowledge argument is one of the most influential.

The first aim of this section is to introduce Jackson's argument. Specifically, I will show the structure of the argument and offer a preliminary clarification of its premises. The remainder of the section will show that this argument appears to target the hypothesis of modest reductionism.

³ Dennett 1991, pp. 399-403; Churchland 1986, pp. 331-334.

⁴ Amongst these, the *Modal Argument*, Kripke 1972, pp. 148-155, the *Explanatory Gap* argument, Levine 1986, Levine 2001 the *Zombies* argument, Block 1978, Chalmers 1996, Chapters 3-4.

Jackson illustrates the knowledge argument with different thought experiments.⁵ The most discussed concerns Mary, a vision scientist:

Who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specialises in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and uses terms like 'red', 'blue' and so on. She discovers for example just which wavelength combinations from the sky stimulate the retina, and exactly how this produces via the central nervous system the contraction of the vocal chords and expulsion of air from the lungs that results in the uttering of the sentence 'The sky is blue'. (Jackson 1982: 471)

Jackson's conclusion follows by answering certain questions about Mary's knowledge after her release.

What will happen when Mary is released from her black and white room or is given a color television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. It is inescapable that her previous knowledge was incomplete. But she had all the physical information. Ergo there is more to have than that, and physicalism is false. (Jackson 1982: 471)

Before considering in detail the argument involved in this thought experiment, a clarification is needed.

Although in the formulation of the argument given above Jackson talks of "physical information", in a subsequent paper he says that Mary in her lab "knows

⁵ In Jackson 1982 and Jackson 1986.

the physical facts about us and our environment”.⁶ Moreover, he claims that the conclusion that physicalism should deny, and presumably that the knowledge argument is implying, is that: “there is more to know than every physical fact”.⁷ Therefore, we can analyse the structure of the knowledge argument in terms of facts.

The knowledge argument moves from the two following epistemic premises about what Mary knows before her release and what she comes to know by seeing coloured objects.

(1) Mary, before her release, has a complete *scientific knowledge* of facts concerning colours and colour vision without having conscious experiences of colours.

(2) Mary, after her release, by seeing a coloured object acquires *new knowledge* about colour experiences.

From these premises, Jackson derives the following ontological conclusion:

(3) There *are facts* that are not physical.

Let us clarify these claims.

In premise (1), Mary's complete scientific knowledge is about the kinds of facts that according to different strands of physicalism exhaust all the facts concerning colour experience. Firstly, Jackson states that:

She knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes everything in completed physics chemistry and neurophysiology. (Jackson 1986: 567)

⁶ Jackson 1986.

⁷ Jackson 1986.

In addition, he claims that Mary knows facts concerning “functional roles” played by the states of the nervous system.⁸ As we saw in the previous chapter, physicalists assume that facts concerning colour experiences belong to these two classes.

The second premise of the knowledge argument concerns knowledge that Mary allegedly acquires by having chromatic colour experiences.⁹ We have seen that he claims that by seeing colours Mary comes to know about the “world and our experience of it”. Here, I will be concentrating on conscious experiences.¹⁰ Jackson claims that this knowledge concerns the occurrence of *qualia*. With the expression “*qualia*”, he intends to refer to:

... certain features of the bodily sensations especially, but also of certain perceptual experiences. (Jackson 1982: 469)

According to Jackson, this suggests that, by seeing coloured objects, Mary comes to know facts about her colour experiences that involve the occurrence of *qualia*. For instance, when Mary sees a red object, she will learn that her experience of a red object has a certain feature. This property is a *quale* that she did not know about before her release. I will refer to this knowledge as *knowledge of what it is like to have a colour experience*.¹¹

The conclusion of the knowledge argument is that there are facts about colour experiences involving the occurrence of non-physical properties. According to

⁸ Jackson 1986, p. 567.

⁹ Jackson assumes that, before her release, Mary can see black and white objects. Presumably, this means that she can also see shades of grey. Thus, she should already know about the *qualia* of the relative achromatic colour experiences. The point of the argument is that by seeing, say, a red object she learns something new about the visual experience of red. On the following, I will omit the specification “chromatic colour”.

¹⁰ Many commentators appear to agree that, according to Jackson, Mary’s supposed new knowledge concerns colour experiences. See, for instance, Carruthers 2000, Papineau 2002, Perry 2001, Levine 2001, Tye 2000.

¹¹ Thomas Nagel introduced the notion of “knowing what it is like” in the philosophical debate in his seminal paper Nagel 1974. Here I use the expression to refer to Mary’s supposed epistemic progress without making assumptions about its existence, nature and content.

Jackson, given that Mary learns about the *qualia* of her colour experiences only upon her release, these properties cannot be physical properties. Otherwise, she would know about them while she was still in her black and white laboratory. Thus, knowing what it is like to have a colour experience is about a non-physical fact concerning the experience that involves the occurrence of a non-physical property or *quale*.

The thought experiment concerning Mary illustrates the knowledge argument as applied to colour experiences. Jackson maintains that similar thought experiments support the conclusion that perceptions in the different modalities and pains have non-physical *qualia*.¹² However, even without such a generalisation, if his argument is sound in the case of colour experiences, the hypothesis that science can accommodate conscious states is undermined. Thus, we can focus our attention on just what the knowledge argument proves about colour experiences. Having illustrated the main structure of the knowledge argument, we have to consider whether it determines a problem for the hypothesis of modest reductionism.

The knowledge argument appears to be directed against a very general formulation of the physicalist thesis about colour experiences. Jackson assumes that Mary knows all the facts of the type studied by physics, chemistry and neuroscience and all the facts concerning causal roles. The argument targets every doctrine based on the thesis that all the facts about colour experiences belong to these two classes. Now, physicalists have formulated this thesis differently. As we saw in the previous chapter, some differences depend upon the ways of understanding the relations between causal roles that define conscious experiences and facts studied by physics and other sciences. It appears that Jackson criticises an assumption shared by these forms of physicalism.

¹² Jackson 1982, pp. 471-472.

Jackson's knowledge argument involves a precise characterisation of physical facts. He denies that his argument requires any definition of "physical information' and the correlative notion of physical property, process, and so on".¹³ Despite this, the argument's logic requires some constraints on the notion of physical fact. From the claim that Mary comes to know a fact she did not know before her release, we have to derive that this is not a physical fact. This appears to require the assumption that *if* a fact is physical, *then* it is potentially knowable or known by Mary when she is still in the black-and-white laboratory.¹⁴ In the following, I will refer to this assumption as the *epistemic constraint*.

The knowledge argument seems to have the resources to target modest reductionism. In fact, this argument addresses those physicalist doctrines that involve a theory-based conception of being physical. The reasons for this are as follows. According to these positions, an entity is physical if it can be contemplated by physical science. Clearly, this implies the epistemic constraint that is central to the knowledge argument.¹⁵ Moreover, modest reductionism involves a theory-based conception of being physical.¹⁶ Thus, it seems that Mary's thought experiment might illustrate a philosophical perplexity about a plausible version of the hypothesis that colour experiences can be studied scientifically.

To sum up, it seems that we have made some progress in our investigation of the problem that conscious experience might create for a scientific account of the mind. The plausible formulation of this hypothesis provided by modest reductionism appears to be targeted by a famous argument proposed by Frank

¹³ Jackson 1982, p. 469.

¹⁴ The importance of this assumption in the knowledge argument is stressed and discussed in Alter 1995, pp. 17-20.

¹⁵ Certain physicalist doctrines can imply the epistemic constraint without recurring to a theory-based conception of being physical. In fact, the epistemic constraint states only a necessary condition for being a physical fact.

¹⁶ See Chapter 1, at p. 10.

Jackson. Nevertheless, such progress might be only apparent. The next section will show that Jackson's argument has to face a serious difficulty.

2.3 Knowledge Beyond our Understanding

Many philosophers have discussed the merits of the knowledge argument by considering what Mary supposedly learns once she leaves the black-and-white laboratory. However, Patricia Churchland and Daniel Dennett have shown that a critical assessment of the argument should start from what Mary knows before her release.¹⁷ Specifically, they have argued that Jackson's assumptions about Mary's scientific knowledge render his argument ineffective.

In this section I illustrate both Dennett and Churchland's objections, and endorse their conclusion. The main point of their criticism is that we lack a clear idea of what Mary's complete scientific knowledge would be. Therefore, we are in no position to establish whether Mary acquires new knowledge about colour experience when she sees a coloured object. Besides, they have argued for the conclusion that we cannot exclude that the knowledge argument might be unsound. I will maintain that this conclusion is plausible given our lack of understanding of Mary's scientific knowledge. However, I will argue that such a conclusion does not receive any independent support from a reformulation of Mary's story provided by Daniel Dennett.

The very possibility of evaluating the knowledge argument appears to be threatened by the assumption that Mary has complete scientific knowledge of physical facts.¹⁸ Clearly, in order to evaluate an argument we should be in the position to understand its premises. However, as Dennett has pointed out, understanding Mary's scientific knowledge is a task:

¹⁷ Dennett 1991, pp. 399-403; Churchland 1986, pp. 331-334.

¹⁸ This interpretation is suggested in Montero 1999.

... so preposterously immense, you can't even try. The crucial premise is that "She has all the *physical* information." That's not readily imaginable.... (Dennett 1991: 399)

In a similar way, Patricia Churchland, who thinks that Mary's complete knowledge might concern the brain, asks:

How can I assess what Mary will know and understand if she knows everything there is to know about the brain? Everything is a lot and it means, in all likelihood, that Mary has a radically different and deeper understanding of the brain than anything barely conceivable in our wildest flights of fancy. (Churchland 1986: 332)

It might be objected that our understanding of Mary's complete scientific knowledge could be based on what we know about contemporary science. However, we cannot be confident that a complete (future or possible) scientific knowledge is going to be similar to the present one in any significant respect that can be intelligible to us. As Patricia Churchland has rightly pointed out, the history of science gives evidence for this.

For to know everything about the brain might well be qualitatively different, and it might be to possess a theory that would permit exactly what the premises says it will not. Utopian neuroscience will probably look as much like existing neuroscience as modern physics looks like Aristotelian physics. So it will not be just more of the same. (Churchland 1986: 332)

An unbridgeable *qualitative* distance might separate contemporary scientific knowledge and Mary's knowledge.

Given the plausible interpretation of Jackson's knowledge argument provided by Patricia Churchland and Daniel Dennett, it emerges that such line of reasoning

cannot be used to evaluate the hypothesis that science can accommodate conscious experiences. In fact, the argument fails to provide an intelligible account of what Mary might know before her release. Moreover, these authors have suggested that Mary might know what it is like to have a colour experience before her release.

Besides promoting an agnostic stance about the plausibility of the knowledge argument, Patricia Churchland and Daniel Dennett have suggested that we cannot exclude possibilities that might render the knowledge argument unsound. Dennett suggests that we cannot exclude a possibility concerning the consequences of possessing complete scientific knowledge. Namely, this knowledge might enable Mary to recognise that the colour experiences she has upon her release fall under the scientific descriptions she already knows. On the other hand, Patricia Churchland maintains that possessing a complete scientific knowledge might *cause* the sort of experience required for knowing what it is like to have a colour experience. Therefore, Jackson's description of Mary's situation before her release might be inconsistent.

Without understanding what Mary knows before her release, many cases are possible. Dennett's criticism of the knowledge argument involves a story about Mary. He appears to think that if before her release Mary lacks knowledge of certain features of colour experiences, then, upon her release, she will not be able to recognise by looking at a blue banana that it has the wrong colour.¹⁹ However, Dennett thinks that the following story shows that this might not be the case:

And so one day, Mary's captors decided it was time for her to see colors. As a trick they prepared a bright blue banana to present as her first color experience ever. Mary took one look at it and said 'Hey! You tried to trick me! Bananas are yellow but this one is blue!' Her captors were dumfounded. How did she do it? 'Simple,' she replied. 'You have

¹⁹ Churchland 1986, p. 333.

to remember that I know everything – absolutely everything – that could ever be known about the causes and effects of color vision. So of course before you brought the banana in, I had already written down, in exquisite detail exactly what a physical impression of a yellow object or a blue object ...would make on my nervous system. So I already knew exactly what thoughts I would have (because, after all, the “mere disposition” to think about this or that is not one of your famous *qualia* is it?). I was not in the slightest surprised by my experience of blue... I realize that it is hard for you to imagine that I could know so much about my reactive dispositions that the way blue affected me came as no surprise. Of course it’s hard for you to imagine. It’s hard for anyone to imagine the consequences of someone knowing absolutely everything about anything!’. (Dennett 1991: 399-400)

Let us evaluate what this story shows about the knowledge argument.

Some maintain that Dennett uses his story to deny that Mary upon release learns anything.²⁰ Nevertheless, his aim in this passage is more modest. As he says:

My point is not that my way of telling the rest of the story proves that Mary *doesn't* learn anything, but that the usual way to imagining the story does not prove that she does. It doesn't prove anything.... (Dennett 1991: 400)

Dennett is not proving that Mary does not undergo an epistemic progress. Rather, he tries to disqualify the intuition that supports the conclusion that she might learn something. To achieve this, he shows that we can imagine a situation in which Mary does not learn anything new by having conscious experiences. In particular, he suggests that Mary's complete knowledge might comprise laws stating causal

²⁰ See Alter 1999 and Chalmers 1996, p. 145. Probably, this reading is suggested by the fact that in other places Dennett has argued for the elimination of *qualia*, see Dennett 1988.

correlations between types of experiences and types of thoughts. In particular, Mary knows that if a certain subject is presented with a blue banana he will have a thought expressible as “this is blue”. Thus by seeing the blue banana she comes to know that people who have this kind of experience think “this is blue”. Therefore, we can have intuitions about Mary’s case that do not support the knowledge argument. However, does this story add anything to the observation that we cannot grasp Mary’s scientific knowledge?

Dennett’s account of Mary’s case exploits the same lack of understanding of her scientific knowledge that he denounces in Jackson’s version. Let us assume, for the sake of the argument, that there can be scientific laws connecting experiences scientifically described and thoughts.²¹ In addition, although Dennett does not explain how this might be the case, we can concede that Mary’s ability to recognise the colour of the object she sees enables her to know what it is like to have a certain colour experience. Still, in order to judge whether Mary passes the blue banana test we need a more substantial understanding of her scientific knowledge.

A central assumption in Dennett's story is that Mary is able to apply the law correlating conscious experiences and the appropriate thoughts just by looking at the banana.²² As Dennett puts it: “Mary took one look at it and said ‘Hey! You tried to trick me!’”. This requires that Mary can recognise the colour experience involved in seeing the blue banana as a certain “physical impression” in her nervous system. However, how does she acquire such a recognitional capacity? It seems that we are not in the position to say. On the other hand, it might be maintained that a relevant ingredient for our coming to decide that Mary is not fooled might be based on our grasp of Mary's notion of “physical impression”. However, our decision cannot be

²¹ Some have argued that thoughts cannot enter into any nomological correlation; see, for example, the very influential Davidson 1970.

²² Howard Robinson shows how this assumption is implicit in Dennett's story, Robinson 1993, p. 175.

based on Dennett's sketchy model of Mary's scientific knowledge. A more substantial account of her knowledge is needed.

We cannot exclude the possibility that just by seeing a coloured object Mary might be able to know that the experience she is having satisfies some scientific description. This simply derives from the fact that we cannot grasp Mary's complete scientific knowledge. Dennett's elaborate story has little to add. Given our ignorance about Mary's scientific knowledge, we cannot exclude that possessing this knowledge might cause in her the relevant recognitional capacity. However, no evidence for the fact that she might have such a capacity derives from Dennett's account. We have the same amount of reason (or lack of it) to accept his story as we do the one told by Jackson. Let us now consider another possibility we cannot exclude about Mary.

Patricia Churchland has argued convincingly that Mary might be able to know what it is like to have a colour experience before her release. For having complete scientific knowledge might produce colour experience or other appropriate mental states supposedly required for the knowledge of what it is like to have a colour experience. In particular, when Mary is still in her laboratory, she might imagine colours with the aid of her scientific knowledge. As pointed out by Patricia Churchland:

For all I know, she might even be able to produce red in her imagination if she knows what brain states are relevant. One cannot be confident that such an exercise of the imagination must be empirically impossible.

(Churchland 1986: 333)

How can we exclude that her complete scientific knowledge might create in Mary, when she is still in the laboratory, the *right mental states* required for knowing what it is like to have a colour experience? This empirical question cannot be solved

before we have such a complete scientific knowledge.²³ It follows that Jackson's account of Mary's situation while she is still in her room might be inconsistent. Mary might not have complete scientific knowledge without having colour experiences and thus knowledge of what it is like to have them.

To conclude, it seems that Jackson's account of Mary's case escapes our understanding. In particular, lacking the knowledge of empirical facts about the development of science, we cannot establish the soundness of his version of the knowledge argument. Jackson's knowledge argument cannot help in our quest for a difficulty that conscious experiences might create for modest reductionism. Does it follow, then, that the *type of strategy* involved in Jackson's argument is of no use for us? The next section will take the first steps towards answering to this question.

2.4 Revising the Knowledge Argument

Articulating a problem for the hypothesis of modest reductionism in terms of Jackson's knowledge argument faces a serious difficulty. Although at first it appeared plausible that this line of reasoning might target such a hypothesis, we now realise that the knowledge argument involves a premise about Mary's knowledge that is beyond our understanding. Therefore, it seems that our initial objective cannot be achieved.

This section suggests a way out of the difficulty. Instead of considering Jackson's version of the knowledge argument, which is aimed at a highly abstract version of physicalism, we should consider whether his thought-experiment could be reformulated to pose a problem for the hypothesis of modest reductionism. It will emerge that pursuing this strategy might be viable only if we provide a more detailed account of the latter hypothesis.

²³ For a similar observation that future (or possible) scientific knowledge might cause knowledge of what it is like to have an experience, see Lewis 1990, p. 580.

A line of reasoning of the type of the knowledge argument might illustrate problems that colour experience raise for modest reductionism. As we have seen, Jackson's version of the argument appears to fail. However, this happens because his argument involves the strong assumption that Mary possesses a complete (future or possible) scientific knowledge of colour and colour experiences. Conversely, the hypothesis of modest reductionism is formulated with reference to a science related to a contemporary physics of ordinary matter. This is the claim that conscious experiences can be described and explained in terms of a science that can be explanatorily interfaced with a theory of the type that currently underlies the study of macroscopic physical phenomena. Thus, if we ascribe to Mary this type of knowledge, we might be able to understand the nature and extent of what she knows before her release. Consequently, we could have an intelligible version of the knowledge argument that might raise a problem for modest reductionism.

My suggestion is to use a suitably modified version of the thought-experiment about Mary to pose a problem for modest reductionism. We are now no longer to assume that Mary knows all the physical facts contemplated by a complete (future or possible) science. Instead, we need to assume that her knowledge is of the kind that the physicalist hope will provide unification with a physical science of ordinary matter. Specifically, this body of knowledge should conform to the requirements of modest reductionism.

Following this suggestion would lead to a version of the knowledge argument weaker than Jackson's. His argument is directed against a general metaphysical version of physicalism, and is meant to prove the existence of non-physical facts and properties. The suggested reformulation of Mary's thought experiment should provide an argument directed against physicalism understood as a methodological strategy. The argument would be used to test the hypothesis that completing a science of colour experience of the type we possess presently, and that can be explanatorily interfaced with physics of ordinary matter, can provide an exhaustive

account of these mental states. Thus, what might be lost in terms of the generality and strength of the conclusion derivable from Mary's case, might be acquired in terms of specificity and intelligibility of the knowledge argument. Nevertheless, are we in the position to formulate such a version of the argument?

The formulation of modest reductionism that has been offered so far is too vague to provide an intelligible version of the knowledge argument. In fact, this formulation does not characterise what Mary might know before her release. On this view, she should possess a scientific account of colour experiences that can be explanatory interfaced with a contemporary physics of macroscopic phenomena. Surely, we can grasp the nature of a contemporary physics of ordinary matter. Nevertheless, we have no indication of what kind of scientific knowledge of colour experiences Mary might possess before her release. This does not mean that we should abandon the project of reformulating the knowledge argument.

Our understanding of Mary's scientific knowledge might be based on a contemporary scientific investigation of colour experiences. It appears worth considering whether currently there is a plausible scientific programme for the description and explanation of colour experiences that can be explanatory interfaced with a current physics of macro-phenomena. Without such an account, we cannot formulate an intelligible version of the knowledge argument.

To conclude, despite the limitations of Jackson's line of reasoning, the knowledge argument's general strategy might be still useful for our project of investigating the tenability of modest reductionism. However, ascertaining whether this is the case requires formulating modest reductionism in more detail.

2.5 Conclusion

We started with the task of formulating the problem that conscious experiences might create for modest reductionism. The knowledge argument is a very influential formulation of a difficulty that these mental states might create for any scientific

account of the mind. However, given a plausible reading of the knowledge argument, this line of reasoning appears to be based on the unintelligible premise that Mary has a complete scientific knowledge of colour and colour vision. Despite this negative conclusion, the next section will attempt to offer an intelligible version of the knowledge argument that targets modest reductionism about colour experiences.

3 Mary's Scientific Knowledge

3.1 Introduction

The previous chapter showed that Frank Jackson's knowledge argument does not challenge the hypothesis of modest reductionism. This failure depends on the knowledge argument's unintelligible premise that Mary has complete scientific knowledge of colours and colour vision.

This chapter aims to present an intelligible argument in the style of the one offered by Jackson that might target a plausible version of the modest reductionism hypothesis. Specifically, I will outline Mary's complete scientific knowledge by referring to contemporary disciplines such as psychophysics, sensory neurophysiology and psychometric.

Section 3.2 illustrates how contemporary science describes colours. Section 3.3 presents the centrality of these descriptions in a contemporary scientific programme for the study of colour experiences. Sections 3.4 and 3.5 show that ascribing to Mary scientific knowledge based on this programme, leads to a version of the knowledge argument that might challenge modest reductionism.

3.2 Colour Spaces

Advancing a version of the knowledge argument that can target modest reductionism requires characterising the type of knowledge that Mary might possess before her release. This characterisation might be achieved by considering a contemporary science of colour vision.

This section illustrates the notion of *colour space* used to describe the colours we discriminate.¹ Specifically, I will present the empirical and computational procedures that lead to the determination of quality spaces. The relevance of these models for the scientific description and explanation of colour experiences will emerge in the next section.

In contemporary science, *spatial models* are used to describe how we categorise different types of stimuli. Here we will be interested in colour spaces that describe our categorisation of colours.² The *colour solid* is an example of colour space. The colour solid represents the ordering of colours that we discriminate by means of three dimensions: *hue*, *saturation* and *lightness* or *brightness* (See Figure 1).

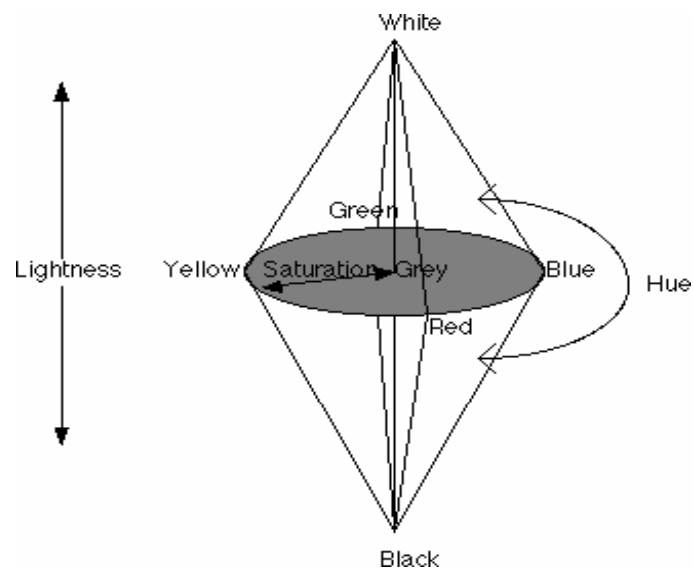


Figure 1 - *The Colour Solid*

Hue is the dimension we normally associate with the basic colours of surfaces. In the colour solid, the hue is represented by the angular direction in the horizontal plane comprised from the central axis of the solid to the position of the point that

¹ The central role of quality spaces in contemporary colour vision science has been explored in detail in Clark 1993 and Clark 2000. See also Palmer 1999a.

² The use of qualitative spaces is not restricted to colour vision. See Clark 2000, Chapter 5, for an illustration of quality spaces that are used to describe other sensory modalities.

represents that colour. Saturation is the “vividness” of a colour. Chromatic colours of the same hue can differ in the strength of the hue. Less saturated colours are closer to grey than the more saturated ones. In the colour solid, the saturation of a certain colour is represented by the distance of the point that represents that colour to the central axis of the colour solid. The third dimension of colour is brightness. Brightness is the relative lightness or darkness of a particular colour, from black to white.³ In the colour solid, the brightness of a certain colour is represented by the height of the point that represents that colour.

The colour solid encodes information about the colours that subjects discriminate. Points in this space are taken to represent colour shades. For these points are individuated by specific values in the axes representing hue, saturation and lightness. In addition, the relative distances of points standing colour shades represent the relative similarities between these shades. For example, orange would be situated between red and yellow, given that subjects find it more similar to these two colours than to, say, blue.⁴ Let us consider how colour spaces are determined.

There are experimental psychophysical procedures to determine how individuals categorise colours. These procedures consist in stimulating the subjects' visual system and registering their discriminatory responses. The stimuli are lights characterised physically in terms of the wavelength and intensity of the electromagnetic waves that compose them. The discriminatory responses are observable behaviours; usually scientists consider verbal reports, but other kinds of

³ Usually brightness is used to indicate a feature of colours seen through apertures or of the colours of self-luminous objects like the sun or lamps. Instead, the term *lightness* is used to refer to a feature of the colours of objects that are not seen through apertures or perceived as self-luminous.

⁴ The colour solid also encodes information about relations of *composition* between colours. Certain hues can be analysed in terms of hues that are more basic. Orange, for instance, appears to contain both redness and yellowness. For this reason, a certain shade of orange will be represented by a location between red and yellow. In contrast, particular shades of red, green, blue and yellow do not appear to be composed of any colours. The colour solid also represents relations of *opponency* between colours. For example, the space shows that there are no hues that appear reddish and greenish. In fact, there is no point in the colour space that might represent those hues.

behavioural clues can be employed. In particular, psychophysicists use a notion of indiscriminability that satisfies some statistically refined conditions. Thus, establishing whether two stimuli are indiscriminable requires reiterated presentations of pairs of stimuli of the same type. For example, if different individuals with normal visual systems fail to notice any difference between two stimuli in any statistically significant way, then the two stimuli are said to be indiscriminable. These procedures determine classes of indiscriminable stimuli.

The colour solid is obtained by mathematical procedures applied to classes of indiscriminable stimuli. These methods have to determine the number of the dimensions of the colour solid and the structure of the relations of similarity between its points. Different methods have been investigated to this end.⁵ A family of statistical procedures, known as *multidimensional scaling* (MDS), has been successfully employed in psychophysics.⁶ Let us illustrate these techniques with an example.

In colour science, a procedure of multidimensional scaling can be applied to similarity matrices representing a relation of similarity between stimuli. For instance, Table 1 shows a matrix representing similarity ratings of monochromatic light stimuli described in terms of their wavelength.⁷ For each pair of stimuli a numeric value represents their degree of similarity. These latter values have been determined experimentally by registering subjects' responses.

⁵ Some philosophers investigated these procedures. Rudolf Carnap, for instance, in his attempt to provide a method for the constitution of all scientific concepts from an observational base, faced the problem of determining colour classes starting from classes of couples of certain primitive particulars, Carnap 1967, pp. 107-136 and pp. 178-182. Nelson Goodman has subsequently shown some limitations in Carnap's methods and has developed an alternative approach. See Chapter 5 of Goodman 1977. For a comparison of these two approaches, see Clark 1993, pp. 101-112.

⁶ The use *MDS* for the determination of the colour space was advocated in Shepard 1962a. An introduction to *MDS* is offered in Clark 1993, pp. 210-221, a more exhaustive and technical presentation can be found in Shifman, Reynolds, and Young 1981.

⁷ Monochromatic stimuli are those characterised by a single wavelength.

<i>Wavelength</i>	445	465	504	537	584	600	651	674
445	-	9	7	6	2	2	7	8
465		-	8	7	2	2	6	7
504			-	9	6	5	2	2
537				-	7	6	3	2
584					-	8	4	3
600						-	5	4
651							-	9
674								-

Table 1 A similarity Matrix for the Observers based on Ekman 1954.

Given matrices of this type, a spatial model is determined by applying the algorithms of multidimensional scaling.⁸ For instance, Figure 2 shows a map obtained by applying *MDS* to the colour similarity data in the Table 1. The *MDS* procedure determines that the similarities between stimuli concern only two dimensions: saturation and hue. This is because the initial matrix is not complete and does not contain enough information to derive the dimension of brightness. However, in principle, if a complete matrix of similarity is available, the structure and dimensionality of complete colour space is obtainable in the same way.

⁸ The procedure here illustrated is an instance of *metric* multidimensional scaling. In fact, the matrix under consideration presents the degree of similarity between the responses evoked by two colour stimuli in terms of a specific numerical value. However, it is possible to generate a colour space by applying *non-metric MDS* procedures. In this case, the matrix does not contain values concerning the degree of similarity between stimuli. Specifically, it is possible to obtain a colour space by applying non-metric *MDS* to matrices representing judgements involving the triadic relation “x is more similar to y than z”. See Clark 1993, pp. 220-221.

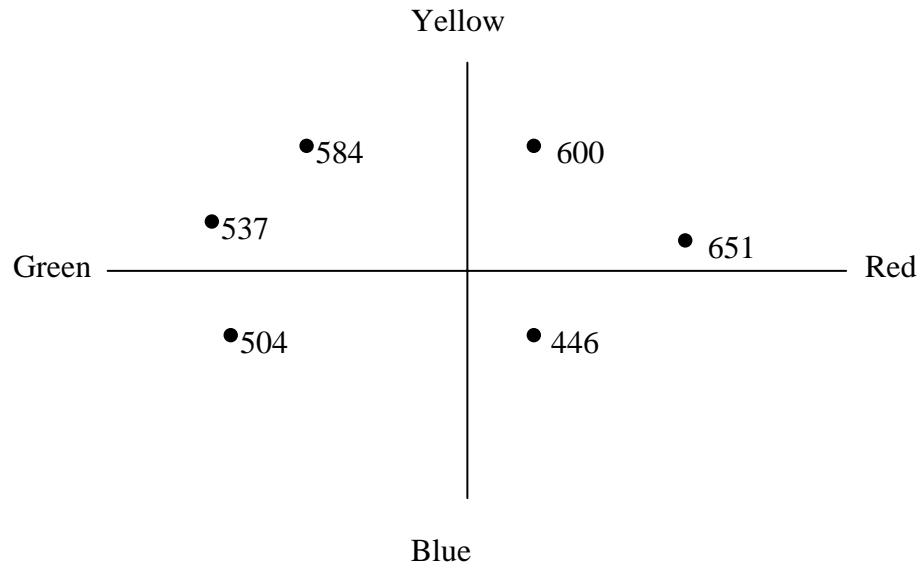


Figure 2 Multidimensional scaling map based on the colour similarity data in Table 1. (Coren, Ward, and Enns 1999: 46)

To sum up, we have seen that, in principle, from classes of discriminatory responses to light stimuli it is possible to determine the dimensions of variations that are involved in our classification of colours. In particular, we have seen that each colour can be individuated and described as a certain position in the colour space. However, we have to determine how contemporary science can describe colour experiences by means of a colour space.

3.3 The Scientific Account of Colour Experiences

Illustrating the colour solid does not appear by itself to advance our understanding of how contemporary science aims to account for colour experiences. Thus, we are left with the problem of providing a substantive account of how Mary might describe and explain colour experiences.

The aim of this section is to address this problem. First, I will illustrate how colour spaces might be taken to provide a description of different types of colour experiences. Second, I will outline the type of scientific explanations as to why we have these different types of colour experiences.

One might maintain that colour spaces describe colour experiences because colours are properties of colour experiences that individuate these mental states. For example, the cognitive scientist Stephen Palmer claims that colour spaces are:

... multidimensional spatial representation (or model) in which different colour experiences correspond to different points in the model. (Palmer 1999a: 924)

Probably, the justification for this view derives from Palmer's position on the nature of colours:

Color is a psychological property of our visual experiences when we look at objects and lights, not a *physical* property of those objects and lights. (Palmer 1999b: 95)

Although this explains how the colour solid describes experiences, nevertheless it involves a very strong claim about the nature of colours.

Such a commitment to a theory of colour should be avoided. Views on the relation between colours and colour experiences vary. Similarly to Palmer, some take colours to be properties of experiences or other mental items.⁹ Others, in contrast, regard colours as properties whose instantiations in an object amount to the power of the object to produce colour experiences of a certain type.¹⁰ Finally, some take colours to be kinds of objective properties of external objects.¹¹ The question is whether there is a way to correlate colour spaces and colour experiences, which is relatively neutral with regard to these theories of colour. Investigating this issue requires introducing some notions.

⁹ Arguments against the identification of colours with physical properties can be found in Hardin 1988.

¹⁰ See McGinn 1991, Peacocke 1984.

¹¹ See Byrne and Hilbert 2003, Tye 2000.

Coloured objects look to subject in certain ways. Let us consider, for instance, the case that a subject sees a red patch. This patch appears to her to have certain features like its shape or colour. In particular, it looks red to her. Seeing an object of a certain colour and seeing an object that merely looks to have that colour involve some similarity. For instance, there are colour illusions caused by changes in illumination or by simultaneous contrast. An example of illusions of the first type is given by seeing a red patch under a yellow light. In this case, the patch visually appears to be blue. The patch looks the same colour to the subject as a blue one. In one standard demonstration of simultaneous contrast, a grey square on a red background appears greenish. The same square seen against a green background appears reddish.¹² In these two cases, we can say that the square looks, respectively, the same colour as a greenish square or as a reddish square.¹³ Therefore, even in cases of illusory perception, the object appears to have certain colour.

The colour space describes the colour that objects look to have. The previous section showed that colour spaces are determined by considering discriminatory responses to colour stimuli. A central assumption in psychophysics is that if certain individuals cannot discriminate two stimuli, then these stimuli will look the same colour to them.¹⁴ Moreover, from matrices that arrange colour stimuli in accordance with the discriminatory responses they elicit, it is possible to determine the dimensions along which these discriminations are performed. For example, in the case of colour solid, these dimensions are saturation, hue, and brightness. Being

¹² See Palmer 1999b.

¹³ Assuming such a similarity does not involve taking a position on the nature of colour. We are not saying whether the colour of an object is the colour it *appears visually* to have in certain condition. In addition, we do not need to assume that colours are properties of colour experiences.

¹⁴ More precisely, it is assumed that if two stimuli are *globally indiscriminable* (by certain individuals in certain conditions) then they look the same colour (to those individuals in those conditions). A certain stimulus *x* is globally indiscriminable from a stimulus *y*, if and only if for any stimulus *z*, *x* is indiscriminable from *z* if and only if *y* is as well. The reasons for adopting the notion of global indiscriminability in order to characterise the relation of looking the same colour are discussed in Clark 1993, pp. 56-62.

derived from discriminatory responses taken to correspond to relations of looking the same colour, these dimensions can be interpreted as dimensions of variations in colour. In particular, specific values in these dimensions individuate a location in the colour space. This location can be regarded as a description of the colour that an object visually appears to have.¹⁵

Besides the colours objects appear to have, we can also assume that there are *qualitative properties* of colour experiences. These latter are properties of colour experiences.¹⁶ Specifically, qualitative properties individuate colour experiences. Thus, if two colour experiences differ, they have different qualitative properties. However, it seems that we can establish an important relation between qualitative properties and phenomenal properties.

Some authors have argued that colour experiences can be categorised in terms of their typical causes.¹⁷ A colour experience is of a certain type, or in our terminology has a certain qualitative property, if a certain paradigm object under certain conditions would produce it. For example, we can say that a *red-type* colour experience is the type produced by a certain paradigmatically red thing in certain suitable circumstances.

It has also been suggested that when something looks to be a certain colour to a subject, either in the perceptual case or in the illusory one, then the subject has a certain type of colour experience. Thus, a subject can undergo a colour experience that is not produced by the paradigm object involved in the characterisation of that type of colour experience. In such cases, that type of colour experience is ascribed

¹⁵ Introducing the notion of a colour that an object visually appears to have does not involve any philosophical stance on this appearance. Whatever this appearance might turn out to be, here I am just showing that we can have a way to provide a description of it by analysing the discriminatory responses of subjects.

¹⁶ This terminology is provided in Strawson 1989. See also Sellars 1963, pp. 93-94, 192-93 and Clark 2000, p. 6.

¹⁷ A typology of this kind is suggested in Peacocke 1984, pp. 349-350, and further elaborated in Millar 1991b, pp. 25-31.

because, to the subject, the stimulus looks as the paradigm object. For example, in the case in which a subject sees a grey patch in a yellow background, we would say that he has the type of experience that a reddish patch would produce in certain circumstances.

The typology I have just described suggests a correspondence between the colour that an object looks to have and the qualitative property that specify a type of colour experience. On this account, if a stimulus looks to a subject the same colour as a certain paradigm stimulus, then such a subject has the type of colour experience that would be produced by the paradigm stimulus. Thus, an individual *S* has a colour experience with a certain qualitative property, when something looks to *S* to have the same colour as the paradigm stimulus.

The descriptions provided by a colour space can be used to categorise colour experiences. In fact, the points in the colour space represent the colours that objects look to have. Thus, to every colour described by the colour space one can associate a corresponding description of a certain type of colour experience. For instance, let us assume that a certain shade is represented by the position *XYZ* in a colour space.¹⁸ The description *XYZ* specifies values of hue, saturation and brightness in terms of a system of relations of similarity with the other colours. The relative type of colour experience will, then, be described as the type of experience someone has when something looks *XYZ* to him.

We have seen how quality spaces can be interpreted as descriptions of the different type of colour experiences. These models provide the *explanandum* for the neuroscience of colour vision. In particular, this suggests a substantive general scientific approach to colour experiences that is consistent with the requirements of

¹⁸ Each term that refers to a shade can be defined by a description that refers only to the relations of similarity (and those of opponency and composition) represented by the colour space. This can be achieved with the logical technique involving "Ramsey sentences". For more details, see Clark 2000, pp. 256-257.

modest reductionism. Thus, we have to determine that such a specific research project, outlined by reflecting on the practices of contemporary colour vision science, can be the target of the knowledge argument.

The final aim of the scientific study of colour vision is to explain the data collected by psychophysics in terms of neurophysiological mechanisms.¹⁹ Although this neuroscientific knowledge is far from complete, we can still be said to have an idea how neuroscience explains the structure of the colour quality space, and thus of the related colour experiences, by considering a widely accepted theory of colour vision.

We have seen that in the case of the colour solid every shade can be identified in terms of its hue, saturation, and brightness. Combinations of values in the dimension of variations of the colour solid can specify the qualitative properties of experiences. In fact, each type of colour experience is defined as the experience that someone has when something appears to him a certain shade of colour. Moreover, the colour solid describes the colours that objects look to have. Thus, the relative ordering of similarity between colours has to be reflected in an ordering of similarity between types of colour experiences.

The structure of the quality space is then explained by finding neurophysiological mechanisms that stand to each other in the same pattern of relations as the points in the quality space. An example of how this might work is provided by a physiological account of the three axes of the colour solid.

According to the *opponent processors* theory, for example, it is possible to explain the structure of the colour solid and thus the location within it of any colour

¹⁹ In the ordinary scientific practice, the phenomena described by psychophysics are often used for postulating the relative neural mechanisms. See on the general logical structure of these inferences, Teller 1984.

shade in terms of the activity of certain neural mechanisms.²⁰ These groups of neurons compare, given their excitatory and inhibitory connections, outputs of three different types of photoreceptors in the retina. These opponent processors have a positive response to stimuli in a certain part of the spectrum and a negative response to those in other parts of the spectrum.

It is assumed on the basis of certain empirical evidence that there are two chromatic opponent processors: blue-yellow and red-green, and an achromatic one: white-black. These three opponent processors generate the three axes of the colour solid. In fact, the different activation of each of these processors, given a certain stimulus, determines the position of the evoked shade with respect of each of the solid's axes. A very bright orange, for example, will result from the combined activation of the neural processes correlated with the experience of red and of yellow and white, and the inhibition of the correlates of green, blue and black.

To conclude, I have provided a model of contemporary scientific study of colour experiences. Moreover, it seems that such a model might satisfy the requirements of modest reductionism. Now, it remains to be considered whether this account can provide a model that can be adapted to Mary's case.

3.4 Mary's Complete Knowledge

In this section, I investigate the plausibility of a version of Mary's thought experiment that ascribes to her a scientific knowledge of the type delineated in the previous sections. Firstly, I analyse the type of understanding that such descriptions provide of colour experiences. Then, I argue that we can reasonably assume that Mary, before her release, has an understanding of this type. So, although we lack all the details involved in the complete colour space that Mary might use to classify

²⁰ Hurvich 1981 is a comprehensive and detailed presentation of this theory by one of its most important advocates. See also De Valois and De Valois 1975, pp. 100-110. Simplified accounts of the theory can be found in Hardin 1988 and Clark 1993.

colour experiences, we nevertheless know the type of description she might use. Moreover, a subject who has never had colour experiences can understand this description.

Now, in order to formulate the knowledge argument, we have to be able to grasp the way in which a scientific account of colour experience, based on contemporary colour science, might develop. If we want to formulate the knowledge argument in terms of our comprehension of contemporary science, then we need to understand some features that define this project. However, these cannot be given in terms of the empirical details that characterise the project. In particular, we cannot say that we already know all the empirical details that would be required for the completion of such research programme. It is enough to consider the determination of the colour solid.

The determination of the colour solid is presently far from complete. One reason is that the set of discriminatory judgements required to achieve this enterprise is very large. For example, if we consider just 20 stimuli, 190 rankings of similarities among pairs are required to fill out the data matrix needed to determine a relative quality space. Each such ranking judgement may require many trials. It is assumed that human subjects can discriminate 10 million of colours, therefore, the determination of only 1 percent of this space would require 5 billion similarity rankings.²¹ This is something that has not yet been achieved and at the present is technically impossible. However, we cannot exclude that a future science might overcome this difficulty. Nevertheless, even if we could overcome these limitations, the colour solid might not be a satisfactory description of colours.

There is no guarantee that a three-dimensional space such as the three-dimensional colour solid would provide an exhaustive description of colour experiences. In fact, a complete colour space would be obtained by considering

²¹ See Clark 1993, p. 118.

discriminatory responses of a limited range of stimuli in very specific experimental conditions. As Austen Clark has pointed out:

... a model that suggests that hue, saturation, and brightness exhaust the dimensions of variations in visual appearances would be true only in a world in which there is one sentient subject, confined to a pitch-black room, allowed to see just one visible point, whose colour qualities are varied in just those three ways. (Clark 2000: 39)

However, in many situations, the colours objects appear to have cannot be completely described in terms of these dimensions of variations.

Glossy surfaces, reflections, translucency, transparency, shadows, and mists all require dimensions of variations in appearance beyond the three sufficient for coloured surfaces or lights presented in the lab. (Clark 2000: 7)

Nevertheless, the general nature and methods that might be used to offer complete description of the colours that objects look to have are clear enough.

The scope of this approach is to determine from a set of discriminatory judgements the sensory dimensions along which subjects can discriminate. Once these dimensions are determined, it is possible to derive a qualitative space. We can assume that Mary knows how colours look to subjects in terms of positions in a completed qualitative space. Such a space models all the possible dimensions of variations amongst colours. These models are obtainable by the application of a multidimensional scaling technique to a complete table of the discriminatory judgements subjects can have about colour stimuli. Surely, we are not able to predict the number of dimensions that will be involved in this space or its structure. Nevertheless, we can assume that Mary will refer to colour experiences by means of relational descriptions modelled by such a qualitative space.

Similarly, the neuroscientific details involved in interfacing explanations of the structure of quality spaces are not complete and the empirical details have to be worked out. However, given the contemporary scientific understanding of the brain, we might expect that these explanations would refer to populations of neurones, and patterns of activation between them, that can be described at the biochemical level by resorting to a contemporary physics of ordinary matter. Thus, we can ascribe to Mary a type of intelligible scientific knowledge that might implement a modest reductive programme. Let us consider whether this ascription can be used to formulate a plausible version of the knowledge argument.

It seems that one of the requirements that the knowledge argument places on Mary's scientific knowledge can be satisfied. Jackson assumes that Mary can have her complete scientific knowledge without having had colour experiences. Thus, we have to establish whether in fact someone who has never had colour experiences can possess the scientific knowledge of the type so far delineated.

The main notions that are involved in the psychophysical categorisation of colour experiences are provided by a colour space that is obtained by certain statistical procedures from discriminatory judgements concerning certain stimuli. It seems that none of these notions requires undergoing colour experiences to be completely understood. Let us consider them one by one.

A colour space is a geometrical representation of the dimensions along which subjects discriminate colours. In the case of the colour solid, we have three dimensions: hue, saturation and brightness. Nevertheless, we can concede that Mary has an n -dimensional model of the different colours. Moreover, she knows the typology of colour experiences derivable from such a colour space. This model will give her information about the dimensions in which light stimuli are categorised by the visual system. Understanding that the visual system enables certain discriminations along these dimensions does not seem to require having colour experiences. Mary will know that these dimensions result from the application of

certain statistical methods to sets of discriminatory responses, or judgements of similarity, elicited by certain physical stimuli. In particular, similar procedures can tell scientists about the dimensions along which a certain species' sensory system, that we do not possess, categorises certain stimuli.

Understanding the notion of discriminatory response involved in psychophysics does not require having colour experience. In fact, Mary understands that these responses are observable behaviours such as a subject's verbal reports about similarities and differences between certain physical stimuli. In particular, she might know that these responses have to satisfy certain statistical conditions in order to count as a reliable registration of the discriminatory capacities of the subjects. Nevertheless, none of these requirements implies that she cannot observe exhaustively these responses in a black-and-white screen.

Finally, we have to see whether Mary can understand the notion of stimulus without having colour experiences. Given the statistical nature of the psychophysical investigation of colour categorisation, stimuli have to be repeatable types. Mary, in keeping with the line of contemporary science, can consider them as classes of presentations that have in common a certain physical property described by the electromagnetic theory of light. In particular, stimuli are characterised in terms of the intensity and wavelength composition of the light that is imaged on the retina. Mary can have an understanding about these features from reading physics books. In addition, she can detect and measure them with instruments, of the type we already possess, the use of which does not require having colour experiences.

To sum up, we can characterise Mary's scientific understanding of colour experiences as related to the ways in which normal individuals discriminate visual physical stimuli, according to the dimensions and the ordering modelled by a complete quality space. We have seen that the concepts involved in this understanding can be possessed without having colour experiences. However, we

now have to consider two types of worries, first encountered in the previous chapter.

3.5 Patricia Churchland and Daniel Dennett Revisited

As we saw in the previous chapter, Patricia Churchland and Daniel Dennett have offered a conclusive criticism to Jackson's knowledge argument. On their view, we cannot grasp what Mary knows before her release. Therefore, the knowledge argument cannot be evaluated. The previous sections showed that a weaker form of the knowledge argument escapes this objection. This version assumes that Mary's complete scientific knowledge is of the *type* involved in contemporary psychophysics and neuroscience. However, these authors have offered other attacks on Jackson's argument that might affect our version of his line of reasoning.

In this section, I will consider these two objections. Patricia Churchland has claimed that Mary's complete scientific knowledge might cause mental states that are required for knowing what it is like to have colour experiences. Dennett has argued that Mary, in having complete scientific knowledge, acquires the ability to recognise the colour experiences she has as falling under certain scientific descriptions she possesses. I will argue that both hypotheses are implausible once we consider the new version of the knowledge argument.

Patricia Churchland maintains that possessing complete scientific knowledge might enable Mary to imagine colour experiences before her release.²² Clearly, such a possibility cannot be excluded *a priori* when we consider a complete colour science without further qualification. However, what about the type of knowledge we have ascribed to Mary by considering contemporary science? It seems that now we are in a better position to answer. Our understanding of Mary's knowledge might help in establishing what she knows before her release. In particular, we have

²² See at p. 49.

to establish whether her knowledge of colour experiences enables her to imagine what it is like to have these mental states.

Our problem has instructive similarities with the following problem discussed by David Hume:

Suppose...a person to have enjoyed his sight thirty years, and to have become perfectly well acquainted with colours of all kinds, excepting one particular shade of blue, for instance, which it never has been his fortune to meet with. Let all the different shades of that colour, except that single one, be placed before him, descending gradually from the deepest to the lightest; 'tis plain, that he will perceived a blank, where that shade is wanting, and will be sensible, that there is a greater distance in that place betwixt the contiguous colours, than in any other. Now I ask, whether 'tis possible for him, from his own imagination, to supply this deficiency, and raise up to himself the idea of that particular shade, tho' it had never been conveyed to him by his senses? I believe there are few but will be of the opinion that he can... (Hume 1978: 6)

Certain commentators think that Hume can conclude that imagination can perform such an extrapolative task because he assumes that colour experiences (*simple impressions* in his terminology) figure in a system of *degrees* of resemblance to each other.²³ In particular it seems that:

Hume also takes it for granted that these internal relations form the linear ordering of a spectrum: each hue occupies a determinate location within a color space composed of distinct (and presumably finitely many) hues. (Fogelin 1984: 267)

²³ Fogelin 1984. Hume's passage concerning the missing shade of blue has been extensively discussed, given that it seems to threaten the consistency of his empiricism about the origins of ideas.

In the case considered by Hume, the subject can imagine the missing shade of blue, because she is in the condition to access certain relational features of similarity between shades she has experienced. These relations provide her a kind of formal understanding of a certain shade in relation to other shades she can see. This allows her imagination to give substance to this formal notion.

It might be maintained that the relational information that Mary possesses, like the information involved in the case of the missing shade of blue, provides knowledge that supports a successful exercise of imagination about an experience of a certain shade. In fact, Mary knows all the relations of similarity between colour shades. However, the two cases present a crucial difference.

In the case of the individual imagining the missing shade of blue, the relations of similarity are between determinate colour shades that she can see. Thus, in addition to understanding the relation of similarity between the shades she can see, she has information about the attributes that ground these similarities. Instead, Mary refers to different types of colour experience with descriptions concerning the colours object look to have. These colours are specified in terms of relations of similarity with other colours. Nevertheless, she cannot access the specific attributes, such hue, saturation and so on whose degree in each colour is expressed by these relations of similarity. In fact, although she has information about relative position in an ordering, she cannot get the absolute value of any of the dimensions (such as hue, saturation, or brightness) that specify each colour. This point can be illustrated with an analogy.

The situation of Mary appears to be similar to that of a person who knows the ordering, and thus the relation of relative similarity, between the heights of the trees in a forest. Although this person has enough information about the relative heights of the trees, he cannot infer on the basis of this information the exact or approximate values of the heights of the trees. Only if he knows the height of a sufficient number of trees (how many depends on the number of trees, the structure

of the order and the approximation desired) will this person be able to infer the height of a certain tree. Having shown that Mary cannot determine, via imagination, what an experience is like, we can turn to the objection offered by Dennett.

Dennett uses the “blue banana trick” thought experiment against Jackson’s knowledge argument.²⁴ As we saw in Chapter 2, this possible case relies on an implicit assumption. Dennett assumes that Mary can recognise the experience she is having of a blue banana as a certain “physical response” that can enable her to recognise that the colour of the object is wrong. The plausibility of this assumption can be evaluated given our understanding of Mary’s scientific knowledge.

Mary knows how blue and yellow things look in terms of the positions of looking yellow and blue in the system of relations of similarity embedded in the complete colour space. However, in seeing the blue banana, Mary can have very limited relational information about the colour of the object. The only relations of similarity and difference she might actually discriminate are those between the way in which the banana looks and the background.²⁵

It could be argued, however, that Mary might know which relational property is involved in the type of experience she is having given her knowledge of its neural correlates. This requires that she can recognise just by seeing the blue banana that she is in a certain brain state. How does she perform this task? Patricia Churchland has maintained that “introspective use of her utopian neuroscience” can help her.²⁶ Unfortunately, the evaluation of a reply that mentions utopian neuroscience is clearly beyond our understanding. Thus, let us consider contemporary neuroscience. It is clear that in order to describe our occurring colour experiences in neuroscientific terms we need a certain learning process. We need to have colour

²⁴ See at p. 50.

²⁵ Churchland 1986, p. 333.

²⁶ Patricia Churchland is endorsing a position defended by Paul Churchland in Churchland 1985, p. 576. This conception is related to a general view on scientific conceptual change expounded in Churchland 1979.

experiences and learn how to describe them in scientific terms. However, Mary, before her release, does not have colour experiences and she cannot undertake this learning process.²⁷

It can be maintained that Mary recognises the type of colour experience she is having, as described by her scientific knowledge, by means other than introspection. In this case, she does not need to have colour experiences.²⁸ However, such assumption creates a problem for Dennett. His argument is based on the assumption that if there is knowledge of what it is like to have a colour experience, then Mary cannot pass the blue banana test. However, even the upholder of the knowledge argument can admit that she can pass such a test when she is still in her laboratory. As pointed out by Dale Jacquette:

The colour scientist can easily know from a third-person perspective that the banana has the wrong natural colour, even while she remains inside her black and white bunker, or indeed, even if she is blind but has access to appropriate Braille readouts from light-monitoring equipment.
(Jacquette 1995: 225)

It is part of her complete knowledge that bananas are yellow, that yellow things have certain physical features or produce certain effects in the visual system. Suppose she sees the blue banana in one of her black and white monitors. Moreover, she can measure the light emitted from it or its effects on her brain. She will, then, conclude that: “this banana has the wrong colour”.

In conclusion, Patricia Churchland and Daniel Dennett's worries do not threaten the consistency of our version of the knowledge argument.

²⁷ A similar reply is in Robinson 1993, p. 175.

²⁸ See on this Jacquette 1995, p. 227.

3.6 Conclusion

In this chapter, two strands of our investigation that concern whether colour experiences raise an insoluble problem for science came together. The first strand derives from the need to offer a plausible formulation of the hypothesis that science can account for colour experiences. I have suggested that, in general, such a hypothesis can be elaborated in accordance with modest reductionism.

The second strand is determined by the need to articulate the philosophical worries about this hypothesis. Many consider Frank Jackson's knowledge argument a plausible way to raise such worries. Relating these two strands has required clarifying Mary's scientific knowledge when she is still in her black and white laboratory.

The result is that intuitions of the type involved in Frank Jackson's knowledge argument might raise a difficulty for modest reductionism. What now remains to be investigated is whether our version of the knowledge argument is successful. Establishing this will require considering in detail the knowledge that Mary supposedly gains by seeing colour experiences. The next chapter will begin to address this issue.

4 Knowing Colour Experiences

4.1 Introduction

The last chapter presented a revised version of the knowledge argument. This line of reasoning appears to raise a problem for a plausible formulation of the hypothesis that science can account for colour experiences. In particular, if Mary's scientific knowledge of colour experiences is of the *type* that we possess presently, upon her release, she will not *recognise* the colour experiences she is having in scientific terms. This claim is weaker than the knowledge argument's conclusion that Mary learns about facts that her scientific knowledge cannot accommodate. Therefore, the plausibility of this conclusion remains to be established.

This chapter examines the thesis that upon her release Mary comes to know what it is like to have colour experiences. Before considering the truth and the implications of this claim, we have to clarify it. Section 4.2 illustrates that this is the thesis that Mary acquires new propositional knowledge about types of colour experiences. Section 4.3 focuses on Jackson's claim that this knowledge concerns the fact that colour experiences have *qualia*. Specifically, such a claim appears in need of justification. The upholder of the knowledge argument should explain how Mary discovers that a certain type of colour experience has a certain *quale*. She might claim that this knowledge derives from the fact that Mary is directly aware of the colour experience and its *quale*. In sections 4.4 and 4.5, I will argue that neither perception nor introspection can deliver such a direct awareness.

4.2 The Fundamental Question about Mary

The central intuition of the knowledge argument is that, by having colour experiences, Mary comes to know something that science can neither describe nor

explain. In particular, Mary comes to know what it is like to have a colour experience.

This section illustrates assumptions that are plausibly involved in the knowledge argument. As a result of this clarification, it will emerge that evaluating this argument requires answering a fundamental question. We have to establish whether, by seeing a coloured object, Mary acquires new propositional knowledge concerning a type of colour experience.

This section clarifies firstly the notion of knowledge of what it is like to have a colour experience that figures in the knowledge argument. It seems that the revised version of the knowledge argument presupposes three requirements for such knowledge. First, knowing what it is like to have a colour experience requires having that experience (or experiences of the appropriate type). Second, this knowledge is propositional. Finally, knowing what is like to have colour experiences should concern *types* of mental states sharable by different individuals. Given this characterisation of knowing what it is like to have a colour experience, the remainder of this section will show which assumptions should be involved in the knowledge argument.

The knowledge argument assumes that one cannot know what it is like to have a certain colour experience without having that colour experience. We can call this the *etiological constraint* on knowing what it is like to have a colour experience. Many commentators on the knowledge argument have recognised this requirement. For instance, David Lewis maintains that the knowledge argument shows that:

Experience is the best teacher in this sense: having an experience is the best way or perhaps the only way, of coming to know what the experience is like. (Lewis 1990: 579)

Similarly, Michael Tye thinks that the knowledge argument points to the problem of the *perspectival subjectivity* of conscious experiences. He states this problem as follows:

What accounts for the fact that fully comprehending the nature of pain, the feeling of depression or the visual experience of red requires having the appropriate experiential perspective (that conferred by being oneself the subject of these or closely related experiences)? (Tye 1995: 15)

Tye's talking of "closely related experiences" invites a further clarification of the etiological thesis.

Besides perceiving coloured objects, Mary might acquire knowledge of what it is like to have a colour experience in other ways. For example, as Jackson has pointed out, false memories might provide this knowledge:

Seeing red and feeling pain impact on us, leaving a memory trace which sustains our knowledge of what it is like to see red and feel pain on the many occasions where we are neither seeing red nor feeling pain. This is why it was always a mistake to say that someone could not know what seeing red and feeling pain is unless they had actually experienced them: false 'memory' traces are enough. (Jackson 1998b: 77)

Similarly, imagining colours we have never seen might ground the knowledge of what is like to have a certain colour experience. In addition, such knowledge might result from odd stimulation. We can see colours by rubbing one's eyelid. Moreover, subjects report that they see colours when they are placed in magnetic fields or are stimulated by electric currents of appropriate voltages flowing through their head.¹ Finally, the direct stimulation of areas of the visual cortex prompts these reports.²

¹ Barlow, Kohn, and Walsh 1947.

² See Penfield 1958.

I will discuss the knowledge argument by focussing the attention on a restricted formulation of the etiological thesis. This claim states that:

- (1) Knowing what it is like to have a colour experience requires seeing coloured objects.

This decision does not appear to prejudge the accuracy of the resulting interpretation of the knowledge argument. In fact, we can safely assume that before her release Mary has never had *any* of the mental states supposedly required for knowing what it is like to have a colour experience. Note that neither the mental states resulting from odd stimulation nor false memories about colours appear to be required to possess scientific knowledge of colours and colour vision.

A second requirement in our version of the knowledge argument is that knowing what it is like to have a colour experience is *propositional knowledge*.³ Before establishing that this is the case, let us clarify this type of knowledge. According to a familiar philosophical view, a central type of knowledge is *propositional* insofar as it involves beliefs construed as attitudes towards propositions. For instance, it is assumed that knowing that the snow is white requires, as one of its necessary conditions, believing the proposition expressed by the sentence “the snow is white”. The proposition is often said to be the *content* of the belief. Another necessary condition for having propositional knowledge is that the belief one has is true. For example, knowing that the snow is white requires believing that the snow is white. Moreover, it requires that the proposition expressed by the sentence “the snow is white” be true. This condition can be spelled out by saying that a proposition is true if and only if a certain fact obtains. In this case, knowing that the snow is white implies that *it is the case* that the snow is

³ Many philosophers think that this assumption should figure also in Jackson’s version of the argument. See Churchland 1989, Lewis 1990, Tye 2000, Perry 2001.

white.⁴ Let us now consider why, when Mary sees a coloured object for the first time, she forms new true beliefs she could not have otherwise formed before her release.

In the knowledge argument, knowing what it is like to have a colour experience must amount to propositional knowledge. This can be argued for from two assumptions. First, Mary's scientific knowledge is propositional. Mary can refer to colour experiences in terms of a description modelled in colour spaces. Now, these descriptions are expressible in linguistic terms. A colour space represents a structure of relations of similarity (composition and opponency) between the ways in which coloured objects look to subjects. Sentences can describe this system of relations.⁵ Second, the logical structure of the knowledge argument requires that the meaning of the term "knowledge" in the two premises is not equivocal. In fact, the knowledge argument can be regarded as involving the following inference:

(1) If it is a physical fact concerning colour vision that *f*, then, before her release, Mary knows that *f*.

(2) Before her release, Mary does not know that that a colour experience of the type of the one she is going to have has a certain *quale*.

Therefore:

(3) The fact that her colour experience has a *quale* is not physical.

Thus, knowing what it is like to have a colour experience must constitute propositional knowledge. Let us now consider another constraint that the knowledge argument puts on Mary's supposed new knowledge.

⁴ The nature of propositions and their relations with facts, linguistic meaning and psychological content are highly debated issues in philosophy. See for an introduction Loux 1998, pp. 133-164.

⁵ Austen Clark illustrates how these descriptions are obtained by means of a logical technique known as "Ramsification", Clark 2000, pp. 255-258.

It might be maintained that the peculiar nature of knowing what it is like to have an experience is related to the *logical privacy* of colour experiences. According to this view, necessarily only the individual who has a colour experience can know about its occurrence and nature. Fred Dretske expresses well the problem generated by this conception:

How can we know what a neighbour's experience - not to mention the experience of aliens and animals - is like unless we can somehow get in their head and experience what they are experiencing? (Dretske 1995: 81)

In particular, if experiences are logically private, scientific knowledge cannot accommodate them. For scientific knowledge is taken to concern facts in principle accessible to different individuals. However, it does not seem that the knowledge argument relies on the privacy of colour experiences.

The knowledge argument is not meant to raise the problem of logical privacy of experiences. Instead, this argument hinges on a comparison between scientific knowledge that can be possessed without having colour experiences and knowledge of colour experiences sharable by those who have these mental states.⁶ As pointed out by Jackson:

The knowledge Mary lacked which is of particular point for the knowledge argument against physicalism is knowledge about the experiences of others, not about her own. [...] The trouble for physicalism is that, after Mary sees her first ripe tomato, she will realize

⁶ In addition, the knowledge argument requires that we understand what a certain hypothetical subject like Mary knows about colour experiences when she has them. The doctrine of logical privacy of mental states denies exactly the possibility of such an understanding.

how impoverished her conception of the mental life of *others* has been *all along*. (Jackson 1986: 567-568)⁷

If Mary upon her release acquires knowledge about colour experiences of others, she needs to think about her colour experiences as *types*. When Mary first sees colours, she discovers that the features of her occurring experiences are features of other people's experiences as well.

Thus, a third constraint on knowing what it is like to have an experience emerges. Knowing what it is like to have colour experiences concerns types of mental states sharable with other individuals. Now, it is worth considering how this requirement rules out an interpretation of Mary's new knowledge.

It might be maintained that what Mary supposedly learns when she sees a coloured object concerns the occurrence of her colour experience. In this case, it is clear that she did not have this knowledge before her release because the occurrence of her experience did not exist. However, this knowledge will not suit the purpose of the knowledge argument. In particular, the physicalist can argue that she comes to know something physical. As pointed out by Jackson:

When she is let out, she has new experiences, color experiences she has never had before. It is not, therefore, an objection to physicalism that she learns something on being let out. Before she was let out, she could not have known facts about her experience of red for there were no such facts to know. That physicalists and nonphysicalists alike can agree on. After she is let out, things change; and physicalism can happily admit that she learn this; after all, some physical things will change, for

⁷ John Perry illustrates this point clearly. He says that Mary can express her discovery as follows: "this is what it is like to see red now, and what it would have been for me to see red before, and what it is and has been and will be for others to see red, in normal conditions with normal eyesight." (Perry 2001: 95).

instance her brain states and their functional roles. (Jackson 1986: 567-568)

We have, then, a reading of Mary's new knowledge compatible with physicalism.

This interpretation misses the point and so does the physicalist's reply that is based on it.⁸ For if Mary learns about the experiences of others, she comes to know about facts that obtained before and after the moment she is having a certain colour experience. In particular, given this requirement, Jackson can claim that these are facts that, if her scientific knowledge about colour experiences were complete, she should have known before her release.

So far, we have seen that knowing what it is like to have a colour experience constitutes propositional knowledge of types of colour experience that Mary can share with other individuals. Let us consider how this characterisation of Mary's supposed new knowledge affects our understanding of the knowledge argument.

The crucial step in the knowledge argument can be described as a passage from an epistemic premise to an ontological conclusion. The epistemic premise is that by seeing colours Mary learns that colour experiences have *qualia*. The ontological conclusion is that there are facts concerning *qualia* that cannot be dealt with by her scientific knowledge.

The main inferential move in the knowledge argument appears to depend on two main assumptions.⁹ The first assumption is that upon her release Mary acquires new propositional knowledge about colour experiences. This implies that she acquires *new* true beliefs she could not have had before her release. In particular, when she sees a coloured object she acquires the new true belief *that* her colour experience has a certain *quale*. The second assumption is that having these new true beliefs implies that there are *new* facts she did not know before her release. Namely,

⁸ A criticism of this interpretation can be found in Lewis 1990, p. 582.

⁹ For a discussion of these two assumptions, see section 3.2.

she comes to know that colour experience have *qualia*. Let us examine this latter assumption.

Inferring the existence of new facts, from the assumption that Mary acquires new knowledge, requires some principles that connects having knowledge to the existence of facts. Such a passage appears to involve some principle that bridges true beliefs and facts. The upholder of the knowledge argument might use two ways to connect true beliefs and facts. First, she might advert to what can be called a *Russellian* account of the content of beliefs.¹⁰ According to this view, the contents of true beliefs should be regarded as collections of actual entities making up *facts* or *states of affairs*. Thus, believing the true belief that “Plato is a philosopher” is being related to an ordered pair composed of the individual Plato and the property of being a philosopher. This account of belief content provides the connection between true beliefs and facts required by the knowledge argument. If Mary has the true belief that her colour experience has a *quale*, it follows that there is the fact that the colour experience has a *quale*.

Alternatively, the upholder of the knowledge argument can support the inference from true beliefs to facts with a *correspondence theory of truth*. Such a theory states that facts are what make beliefs true.¹¹ The main difference with the previous strategy is that, in this case, facts are thought of as “mirror images” of true proposition involved in true beliefs. Therefore, if Mary acquires a new true belief about what is like to have a colour experience, then there is the fact that renders her belief true.

In particular, it appears reasonable to assume that in the knowledge argument the notion of correspondence between beliefs and facts that makes these belief true is characterised as a relation between structured entities. The ontological conclusion

¹⁰ Russell 1912, Chapter 12.

¹¹ See, for an introductory presentation of this doctrine, Pitcher 1964, and Kirkham 1992, Chapter 4.

of the knowledge argument appears to suggest that the fact concerning what it is like to have an experience is a complex structured entity involving the experience and a certain *quale*. Now, this can be derived by if this fact mirrors the structure of Mary's belief that this fact renders true. This belief can be taken to be composed of the concept [experience] and [*quale*], the experience and the *quale* to which these concepts refer compose the corresponding fact.¹²

It seems that we are now in the position to realise what is in effect the central question underpinning the evaluation of the knowledge argument. Namely, whether by seeing a coloured object, Mary acquires new propositional knowledge about her colour experience. We have illustrated how the supporter of the knowledge argument might justify the truth of the conditional such that: if Mary acquires knowledge of what it is like to have a colour experience, then there is a fact that renders true the belief that is involved in this knowledge. Clearly, we have to investigate whether he can support the antecedent of this conditional.

To recapitulate, I have illustrated certain assumptions concerning the knowledge of what it is like to have a colour experience. Moreover, I have considered the main inferential step in the knowledge argument. This clarification has shown that the main question that has to be investigated is whether Mary, by

¹² This connection could be spelled out in more detail in a formulation of a correspondence theory of truth that regards the relation of correspondence as a form of *isomorphism*. An isomorphism is a function between two structures that preserves the relations between their elements. Given the belief that P (a) its content that P (a) can be regarded as a structure involving the concept [P] and the concept [a] and a certain syntactical relation. Now, let us consider a function *r* that can be regarded as an assignment of reference. The arguments of *r* are the components of the content of the belief, and its values, the components of a fact; *r*(P) is a certain property and *r*(a) is a certain thing. Now, in a version of the correspondence theory of truth, P(a) is true if and only if *r* (Pr(a)). P (a) is a relation between concepts (or words) and *r*(Pr(a)) is an ontological relation, for example instantiation, between a property and a thing. Clearly, when P (a) is true the function *r* maps related constituents of the content of the belief into related constituents of the fact that renders the belief true. Therefore, the function *r* is an isomorphism. See on this version of the correspondence theory of truth, Kirkham 1992, Chapter 4.

seeing coloured objects, acquires new propositional knowledge concerning types of colour experiences that she could not have had before her release.

4.3 A Problem Concerning the Content of Mary's Belief

Assessing whether Mary comes to know what it is like to have colour experiences upon her release requires clarifying the content of the belief involved in this knowledge.

This section investigates Jackson's claim that Mary comes to know that the colour experience she is having has a certain *quale*. In particular, I will consider the account of how Mary acquires this belief. My discussion of this account will be divided in two stages. First, I will argue that such an account is not detailed enough to avoid an ambiguity in the interpretation of the content of Mary's new belief. Second, I will illustrate the kind of assumptions that might be used by the upholder of the knowledge argument to dispel this ambiguity.

The knowledge argument involves the thesis that by seeing a coloured object Mary acquires the new belief that the type of colour experience she is having has a certain feature. In fact, her supposed new knowledge of what it is like to have a colour experience is taken to be propositional. Propositional knowledge is usually regarded as involving beliefs that something is the case. Specifically, Jackson and many who have commented on his argument maintain that Mary's new belief concerns the fact that colour experiences of the same type of the colour experience that she is undergoing have a certain *quale*. For instance, if she sees the sky, she might express her new belief as:

- (1) The experience involved in seeing blue has *this* property *Q*.

In (1) "the experience involved in seeing blue" stands for a type of colour experience, and "this property *Q*" refers to its *quale* that contributes to what it is like to have that type of colour experience.

Ascribing to Mary the belief (1) appears to require two main explanations. First, it has to be explained how Mary can recognise that she is having a colour experience of a certain type. Providing this account is not difficult. Mary knows before her release that seeing a tomato in certain conditions involves the experience of seeing a red object.¹³ Thus, once she recognises that she is seeing a tomato in certain conditions she can form the belief that she is having a colour experience of a certain type. Alternatively, someone can tell her that she is having a certain type of experience. Both suggestions are consistent with the knowledge argument. However, a second and more important explanation has to be given. It has to be explained how she might form the belief that experience of this type has a certain *quale*.

In the knowledge argument, the assumption that Mary comes to believe that colour experience of a certain type has a *quale* appears to be related to the fact that she sees a coloured object. Thus, the following might be the implicit account of how she acquires her supposed new belief:

(2) Mary comes to believe that the colour experience of seeing blue has property *Q* by seeing a blue object *r*.

Therefore, claim (2) should have a role in justifying the idea that the content of Mary's belief about colour experience is expressed by (1). Thus, there should be some kind of transitions from seeing a coloured object to Mary's beliefs about *qualia*, understood as properties of colour experiences.

The account (2) appears to be threatened once we realise that there are two different interpretations of what *qualia* might be. Philosophers have proposed

¹³ This account is suggested in Perry 2001, p. 96.

different notions of *qualia*.¹⁴ On the one hand, *qualia* can be taken to be features of experiences. On the other hand, *qualia* might be regarded as properties that the objects of experience look to have. A passage by Daniel Dennett illustrates, perhaps unintentionally, these two senses:

'*Qualia*' is an unfamiliar term for something that could not be more familiar to each of us: the ways things seem to us ... Look at a glass of milk at sunset; *the way it looks to you* - the particular, personal, subjective visual quality of the glass of milk is the *quale* of your visual experience at the moment. *The way the milk tastes to you then* is another, gustatory *quale*, and *how it sounds to you* as you swallow is an auditory *quale*. These various 'properties of conscious experience' are prime examples of *qualia*. (Dennett 1988: 619)

At the beginning of the passage, *qualia* are "the ways things seem to us". The milk looks sounds or tastes a certain way. In the last line, Dennett says that *qualia* are "properties of conscious experiences". Now, when Mary sees an object it seems intuitively plausible to say that the object will look a certain way to her. Why should we then assume that she comes to know about properties of her experience? It seems that the upholders of the knowledge argument need to support claim (2) by providing a more accurate account of how seeing coloured objects renders available beliefs about *qualia* understood a properties of experiences. To explore how they can achieve this we need some preliminary clarifications.

¹⁴ An influential discussion of these two different senses in which we can talk about the qualitative features of experience is offered in Sellars 1963, pp. 93-94, pp. 192-3. On how the equivocation of these two readings affects some contemporary discussion on the nature of experience see Martin 1998. In recent years, many philosophers have endorsed a representationalist doctrine of colour experience. One of the tenets of representationalism is that the only features we come to know by having experiences are the properties that the objects of perceptual experience are represented to have; see for instance Harman 1990, Dretske 1995, Tye 1995, Tye 2000.

According to statement (2), Mary's supposed new beliefs appear to *be based* on the awareness of something that she acquires in virtue of seeing a coloured object. This notion of beliefs based on awareness can be illustrated in the case of visual awareness. We can distinguish between beliefs that depend on awareness of something and those that do not. My judging that there is a computer screen in front of me might depend on my being visually aware of the computer and its position. If I am blind, I can still believe that there is a computer screen in front of me. Nevertheless, in this case the belief does not depend on my visual awareness of it. Clearly, thesis (2) requires that Mary's supposed new belief depends on what she sees. Let us consider how this idea of dependence can be made more precise.

It is plausible to assume that Mary's belief is *directly based* on the awareness she has in virtue of seeing a coloured object. The difference between beliefs directly and indirectly based on awareness can be illustrated with an example. I can believe that someone is at the door, by seeing him and thus by being visually aware of him. This appears to be a belief directly based on awareness. This can be contrasted with the case in which I believe that someone is at the door when I am aware that the bell is ringing. In this case, the belief is not directly based on my visual awareness of the person at the door. In particular, to form that belief I need to know certain facts about the function of doorbells and that this bell is not malfunctioning. Now, thesis (2) does not explain in detail how Mary forms her supposed new belief. Therefore, we are entitled to explore whether this belief is directly based on her awareness that derives from seeing a coloured object.

We can be aware of different types of entities. First, we can be aware of objects. For example, when we see a cat we are visually aware of an object. Second, we can be aware of properties. Thus, when we see a furry cat we can be aware of the

property of being furry. These two types of perceptual awareness can be indicated as *o-awareness* and *p-awareness* respectively.¹⁵

Beliefs can depend in different ways on what we are aware of. Let us consider the belief that that an object *o* has a certain property *P*. Having such a belief might require both awareness of the object *o* and of its property *P*.¹⁶ For example, if I see a furry cat, my belief that the cat is furry might depend on the fact that I am *o*-aware of the cat I see, and *p-aware* of its property of being furry. However, someone can believe that a certain object *o* has a certain feature *P*, by being just aware of *o* or of *P*. As an example of the first case, consider someone who by seeing a table with the naked eye comes to believe that “this table is composed of atoms”. While the person is visually aware of the table, she is not visually aware of its being composed of atoms. As an example of the second case, consider someone who sees a certain coloured patch and believes “John’s car is this colour”. Although the person is not aware of John’s car, she can have a belief about this car’s colour by being aware of another object’s colour.

There is an intuitive difference between direct and indirect *o*-awareness. For instance, the belief that “The building is new” can be taken to depend on our visual awareness of the building. However, we might say that our awareness concerns the part of the building we are seeing. Thus, our awareness of the building is indirect and depends on the awareness of its part. With these preliminary remarks in place, we can turn to the problem of interpreting Mary’s supposed new belief.

It is plausible to maintain that the upholder of the knowledge argument might endorse what can be called the thesis of the *direct awareness of qualia*. This thesis states that:

¹⁵ These different notions of awareness are spelled out in Dretske 1999.

¹⁶ This idea can be regarded as a generalisation of a thesis concerning visual awareness. In certain circumstances, to be visually aware *that* something is the case we have to see things and their properties. For arguments in support of this latter thesis, see Jackson 1977, pp. 153-164.

(3) Mary comes to believe *that* experience *e* has *quale* *Q* because, in virtue of seeing coloured object *r*, she is directly aware of *Q*.

In fact, the knowledge argument does not contain any detailed account of how Mary acquires her new belief. Claim (3) appears to formulate such an account in the most direct way. In this case, the reticence of the upholder of the knowledge argument might appear to be justified. We do not need to be told anything about how Mary forms her new belief about her colour experience. In fact, this belief is based on a transition as direct as the one involved in forming a belief about a certain object that we perceive. However, it seems that account (3) is in need of further improvement.

To endorse (3), the upholder of the knowledge argument should assume that Mary becomes directly aware of the colour experience she is having. This claim can be supported as follows. When we judge, in virtue of what we see, that something has a certain property, we cannot be directly aware of that property without being equally aware of the object that has that property. For instance, in the case of vision, in order to be visually aware of the circular shape we need to see the circular thing. Therefore, if Mary becomes directly aware of a property of her experience she needs to be aware of the experience in which it is instantiated.

It might be objected that in many cases we believe that something has a certain property, by being aware of that property, without, at the same time, being aware of the object in question. For example, if I believe that John and Fred own cars of the same colour, then I can believe that John's car is a certain colour in virtue of being aware of the colour of Fred's car. Thus, it might be maintained that Mary might be aware of a property of her experience without being aware of the experience. Instead, she is aware of something else that has that property.¹⁷

¹⁷ Some argue for the claim that we can be aware of a property without being aware of any object. See Dretske 1999.

There is a reply to this objection. In fact, in the case of the cars it is presupposed that I know that colours are properties of cars. However, in Mary's case we want to explain how on her release she comes to know about *qualia* as properties of her colour experiences. Thus, if we assume that Mary is not aware of her colour experience, it would remain to be explained how she comes to know that the *quale* is a feature of her colour experience.

To recapitulate, when facing the claim that Mary comes to know that colour experiences have certain *qualia* we cannot ignore the fact that there are different ways of understanding what these "*qualia*" properties might be. In particular, some regard *qualia* as properties of experience and others as properties of their objects. I have shown that the knowledge argument in itself cannot support one reading or the other. However, the upholder of the knowledge argument has a way to justify the claim that Mary acquires new beliefs about the *qualia* of her experiences. In the next section, I will investigate whether we should endorse this explanation.

4.4 Direct Awareness of Experiences and Perception

In the previous section, I considered how the upholder of the knowledge argument might defend the idea that Mary comes to believe that experiences have *qualia*. He might assume that by seeing coloured objects she becomes directly aware of her experiences and their properties. Given this awareness, she can then judge that her experience has a certain *quale*.

This section aims to show that this account is not tenable if we assume that perceiving coloured objects can provide such awareness. In fact, the principal philosophical accounts of perception do not support the idea that perception provides awareness of experiences. Moreover, the only account of perception that supports this idea is untenable.

As we saw in the previous section, the upholder of the knowledge argument might endorse the following account of how Mary acquires her new belief about colour experiences.

(1) Mary comes to believe *that* experience *e* has *quale* *Q* because, in virtue of seeing coloured object *r*, she is directly o-aware of *e* and she is directly p-aware of *Q*.

One plausible interpretation of this account is that it is *solely* in virtue of the perception of a coloured object that we are directly aware of our experiences. Specifically, the upholder of the knowledge argument might argue that the previous interpretation of (1) might be based on a certain philosophical account of perception. Therefore, we have to consider which account of perception she might endorse.

It is not open to the upholder of the knowledge argument to defend (1) by endorsing *direct realism*. This is the doctrine that perception provides direct awareness of physical objects (or their parts) that exists independently of our mind. This excludes the possibility that in perception we can be at the same time directly aware of our experiences. However, the upholder of the knowledge argument can protest that direct realism is not the only account of perception available.

It might be assumed that claim (1) derives from a *sense datum* account of perception. This family of theories, traditionally opposed to direct realism, can be characterised by two main theses. The first one is a negative, revisionary, claim against direct realism. In perceiving, we are not directly aware of external objects that exist independently of our mind. Different arguments have been offered to support this conclusion; the resulting debate is lengthy and still lively.¹⁸ However, another central tenet of *sense datum* theory is relevant for our present concern. The

¹⁸ Howard Robinson presents these arguments, the relative debate, and a defence of a sense-datum theory of perception in Robinson 1994.

positive thesis often advanced by sense datum theorists is that in our perception we are directly aware of sense data. Therefore, the upholder of the knowledge argument might maintain that colour experiences are identical to sense data. Nevertheless, the latter identification is problematic.

The identification of colour experiences with sense data is not consistent with the traditional formulation of sense datum theory of perception. The promoters of this doctrine have argued for a distinction between the act of sensing, the experience, and the thing directly sensed, the sense datum. For example, G. E. Moore, surely one of the main proponents of sense datum theory, maintains that:

I shall always talk of sense-data, when what I mean is such things as this colour, size, and shape, which I actually see. And when I want to talk of my seeing of them I shall expressly call this the seeing of sense-data; or, if I want a term which will apply equally to all the senses I shall speak *of the direct apprehension* of sense-data. (Moore 1953: 49)

Therefore, identifying experiences with sense data would be committing a fallacy of composition. This sort of mistake stems from the assumption that a part is identical to the whole.

Another option available to the upholder of the knowledge argument is to maintain that, in perception, we are directly aware of our colour experience. However, it seems that this proposal can be rejected from an ordinary phenomenological point of view. This type of phenomenological observation can be found in authors that have endorsed very different accounts of experience.¹⁹ Leaving aside their positive account of what we are directly aware of when having colour experiences, these authors have pointed out that we are not directly aware of the colour experiences involved in seeing coloured objects. Instead, it appears that

¹⁹ Moore 1903, Tye 2000, Harman 1990.

we are directly aware of these objects. If we see a red rose, I am aware of the way in which the rose looks to me. However, what we perceive does not make us aware of our experiences. In analogy with a glass polished perfectly, it seems that our ordinary way of considering perception shows that our experiences are *transparent*. We do not “see”, or otherwise become aware of, our experiences just in virtue of perceiving physical objects.

The upholder of the knowledge argument might support claim (1) by endorsing an *adverbial account* of perception. Some philosophers, not sympathetic to sense data, have maintained that what primarily leads us to posit these entities are erroneous ways of understanding certain sentences used to ascribe mental states such as illusions, afterimages and pains. The grammatical form of these statements might suggest that there are objects of experience. These philosophers maintain that, in reality, such expressions concern ways in which people experience. For example, let us consider the grammatical form of “John hallucinates a red circle”. This claim suggests that there is an object that is circular, red and to which John is related by his hallucination. The suggested reconstruction of this ascription is “John hallucinates redly-circularly” (or in a red and circular manner). This adverbial analysis involves the predication of a property to John and no existential quantification over the object of his hallucination.²⁰ The upholder of the knowledge argument might suggest that a proper rendering of the ascription “S is seeing a red thing” is that “S is seeing red-thingly”. However, this seems not to help in supporting the account expressed in (1).

An adverbial account of experience does not support the conclusion that seeing a red object delivers direct awareness of the experience. According to adverbialism, when we see a colour, we sense in a certain way. Therefore, the experience does not

²⁰ Here I am taking adverbs to be operators that, when applied to predicates, determine other predicates. Moreover, I assume that predicates are linguistic counterparts of properties. Many adverbial theorists assume that adverbs are predicates that apply to events. See, Tye 1984, p. 201.

put us in relation with a certain object. *A fortiori*, having the experience cannot be sufficient for having direct o-awareness of itself as required by account (1). Of course, this does not exclude that an adverbial account of experience might be consistent with the possibility of having direct awareness of experiences. However, such awareness has to be delivered by a different faculty from those directly involved in perception. For instance, introspection might have such a role.

To sum up, we have seen that the upholder of the knowledge argument must account for how Mary comes to know that experiences have *qualia*. One proposal might be based on the assumption that she is directly o-aware of her experiences. I have argued that such awareness cannot be derived solely from seeing coloured objects. However, perception does not exhaust all the modes in which we might come to know about our experiences. The next section will consider whether introspection can deliver direct awareness of our experiences.

4.5 Introspection and the Direct Awareness of Experiences

The upholder of the knowledge argument cannot claim that perception provides direct awareness of colour experiences. Given that introspection has been traditionally considered the principal source for knowing our mental states, it might be maintained that Mary forms the belief that colour experiences have certain *qualia* in virtue of such a faculty. In particular, the upholder of the knowledge argument might claim that by seeing coloured objects and then introspecting she becomes directly o-aware of her experiences.

The present section shows that the upholder of the knowledge argument cannot maintain that introspection requires (or consists in) being directly aware of colour experiences. I will argue that central intuitions exploited in the knowledge argument are not consistent with this model of introspection. To support this claim, I will proceed as follows.

First, I draw a distinction between the objects of experience and the acts of experiencing. In particular, the experiencing involved in having a colour experience is an act of *p*-awareness of the colour that an object appears to have. Colour experiences relate us to the colours that the objects of experience have or look to have. Second, I argue that the assumption that Mary is directly *o*-aware of her experiences requires an explanation of how she picks out her colour experience. Finally, I will argue that the upholder of the knowledge argument cannot provide such an account.

Philosophers and psychologists have provided different models of the knowledge of mental items such as experiences, beliefs and the self.²¹ Here I am considering whether there is an account of introspection that can be used by those sympathetic to the knowledge argument to maintain that Mary learns that experiences have certain *qualia*. This is the problem of investigating whether the following assumption is true:

- (1) Mary comes to believe *that* experience *e* has *quale* *Q* because she is *directly o-aware* of *e* and *p-aware* of *Q* in virtue of her introspective awareness of what is going on while she sees a coloured object *r*.

Let us consider a criticism of this account of how Mary can form her supposed new belief about her colour experience.

Many philosophers would maintain that the account (1) is inconsistent with the deliverance of introspective reflection. They have claimed that, in trying to achieve introspective knowledge of our perceptual experiences, we can only be aware of the objects of these experiences and the properties of these objects.²² In particular, they

²¹ A survey of these accounts can be found in the first part of Lyons 1988.

²² Famously G. E. Moore (Moore 1903, p. 25) used similar observations to dismiss the idealists' claim that we are only aware of mental states. Similar observations against the idea that we are aware of our experiences (or their properties) can be found in contemporary advocates of the representationalist theories of the mind, see Tye 1992, p.160 or Harman 1990, p. 667.

have thought that ordinary phenomenological observations support their claim. For instance, let us assume that I am seeing a red patch in front of me and I intend to determine the features of the experience I am having of it. These philosophers maintain that, whatever I might end up thinking or knowing about the colour experience I am having; my direct awareness can only concern the red patch and its features. Therefore, introspective knowledge requires the same or is nothing over and above the awareness involved in perception. The previous conclusion, in conjunction with the plausible assumption, discussed in the previous section - that perception involves awareness of the objects of experiences - implies also that in the introspective case, we are not directly aware of our experiences.

One objection to the previous argument is that by appealing to introspective evidence we are in fact committing an error in philosophical method. Ned Block has recently denounced as inappropriate the use of introspective reflection in philosophy:

When I look at my blue wall, I think that in addition to being aware of the color I can also make myself aware of what it is like to be aware of the color. I am sure others will disagree. The one thing we should agree on is that this is no way to do philosophy. [...] Looking at a blue wall is easy to do, but it is not easy (perhaps not possible) to answer on the basis of introspection alone the highly theoretical question of whether in doing so I am aware of intrinsic properties of the experiences. (Block 1990: 73)

Block is criticising those who claim that introspection shows that properties of experience are representational or extrinsic properties of experience.²³ However, it seems that the appeal to introspection is equally methodologically suspicious in the

²³ In this passage Block is considering the position advanced in Harman 1990.

case here under discussion. What is preliminarily required is some clarification of two notions. First, we have to spell out clearly the notion of direct o-awareness. Second, we need a clear idea of what a colour experience might be. Let us consider the first issue.

By adapting a characterisation of direct perception offered by Paul Snowdon, we can characterise direct awareness.²⁴ According to him, the direct perception of an object appears related to the ability to have certain true demonstrative thoughts about the object that are *independent of* other demonstrative thoughts. For example, if a subject directly perceives a dog in front of her, then the perceptual contact with the dog allows that person to judge that “this is a dog”.²⁵ In addition, such a thought should not depend on any other demonstrative thought.²⁶ This rules out that we have direct perception of an object in cases such as the following. For example, in a case of *deferred ostension*, I can say “this is my dog” by seeing a picture of my dog. However, it seems that in this case the contact with the dog is indirect. In particular, my demonstrative thought about the dog presupposes the truth of the more basic demonstrative thought “This is a picture of my dog”. Thus, we can assume that direct awareness can be characterised as follows:

S is directly o-aware of *y* if and only if *S* stands, in virtue of *S*'s awareness, in such a relation to *y* that, if *S* could make demonstrative judgements, then it would be possible for *S* to make the *true* and independent demonstrative judgement 'That is *y*'.

²⁴ Snowdon 1992. Whether we perceive directly objective entities or mental ones is a central question in philosophy of perception. Clearly investigating this question requires a preliminary clarification of the notion of direct perception. For a discussion of the principal way to spell out this notion and a positive account alternative to that of Snowdon, see Jackson 1977, Chapter 1.

²⁵ This example uses a demonstrative that involves the concept [dog]. There is no need to assume that every case of direct awareness involves the use of a specific concept. The person might be directly aware of a certain object, even when her successful demonstrative reference involves a generic concept such as [object].

²⁶ This condition of independence is illustrated in Snowdon 1992, pp. 58-59.

We have clarified what is required for being directly aware of something. Now, we have to account for the nature of colour experiences.

Many philosophers have regarded colour experiences as internal states that place us in some kind of relation with certain entities.²⁷ In particular, it might be suggested that in a colour experience we can distinguish between the act of awareness that grounds this relation and the property to which we are related.²⁸ For example, if I perceive a red apple, my colour experience relates me to the colour the apple looks to me to have. On the other hand, if I perceive a yellow lemon I am related by the colour experience to the colour the lemon looks to have. Thus we can say that experience involve an act of p-awareness, which enters into every colour experience, from the specific colour we are aware of.

To be directly aware of her colour experience, Mary needs to identify her experience. In fact being directly aware of an object would require being in the position to have a demonstrative thought about that object. As many philosophers have pointed out, this in turn requires that the subject should be able to collect “identification information” about that object.²⁹ This can be illustrated in the perceptual case. When we refer with a demonstrative to an object we perceive, we are capable of distinguishing it from other objects simultaneously present in a certain space or from a background. The subject should possess the discriminatory capacities to pick out the object from other objects. In particular, it seems that these abilities require discriminating features of the objects demonstrated. Moreover, the

²⁷ See the quote by E. G. Moore, reported at p. 98. The classical defence of such a distinction is given in Moore 1903. Theories of experience containing this distinction differ, principally, in dealing with two issues. First, there are dissimilar ways to explain how experiences can makes us aware of something. Second, there are different assumptions on the nature of the entity to which we are related. The main difference is between those that assume that we are aware of external physical objects and those that think that our awareness concerns mental entities such as sense-data.

²⁸ This analysis is suggested in Dretske 1999.

²⁹ The necessity of such a perceptual contribution to successful demonstrative reference has been defended in Evans 1982, p. 107, and p. 149. See also Clark 2000, pp. 131-136, and Millikan 1990.

subject should be able to discriminate relations that hold between the demonstrated objects and other objects or the background. The distinction of the two components in colour experiences might suggest to the upholder of the knowledge argument an account of how Mary can pick out her colour experience.

It might be maintained that it is the act of awareness involved in Mary's colour experience that grounds her ability to pick out the experience she is having. Then Mary's introspective direct o-awareness of her colour experience does not require her to attend to the properties of which this experience makes her aware. Therefore, Mary can have the demonstrative thought about her experience by noticing that she is undergoing an act of awareness.

To this it can be objected that Mary cannot determine the act of awareness involved in her experience without noticing the colour that the object of this experience looks to have. But it might be replied that the knowledge argument might be taken to show that it is awareness, as involved in all colour experiences, that has features that science cannot explain. After all, the knowledge argument is taken to show the limits of our scientific understanding of consciousness. Thus, it can be maintained that the conclusion of the knowledge argument concerns the phenomenon of being aware. In this case, what renders consciousness scientifically intractable are certain properties of awareness.

Perhaps, the nature and workings of awareness are serious problems for science. However, the knowledge argument cannot raise these difficulties. Presumably, by reading her black and white books or watching her instruments, Mary has many different visual experiences. For example, she can experience a shape or a grey surface. Thus, she might attend to the acts of awareness involved in these experiences independently of the awareness of the properties of which she is p-aware in having these experiences. However, then, how can this knowledge differ from knowing what it is like to have a colour experience? It seems that we cannot

point to any difference. Therefore, we might be entitled to conclude that, before her release, Mary knows also what it is like to have colour experiences.

To sum up, Mary cannot become o-aware of her colour experiences without being aware of the properties of which she is aware in virtue of having such experiences. Thus, we have to consider whether another possibility is open to the upholder of the knowledge argument. I have shown that the upholder of the knowledge argument cannot maintain that Mary comes to know that her colour experiences have *qualia* by being directly aware of them in virtue of introspection. Of course, there might be different ways to come to know about our experiences. Moreover, these models might be compatible with the intuitions involved in the knowledge argument.

4.6 Conclusion

This chapter began by examining what is involved in the acquisition of the knowledge of what it is like to have a colour experience. The knowledge argument puts precise constraints on this knowledge. Namely, it is assumed that by seeing coloured objects Mary acquires propositional knowledge. This is the knowledge that colour experiences have *qualia*.

Specifically, I have investigated how the upholder of the knowledge argument can support the assumption that upon her release Mary acquires beliefs that colour experiences have *qualia*. I have argued that such an assumption cannot be based on the thesis that she has direct awareness of her experiences and their features.

In the next chapter, I will consider whether there are other accounts that the upholder of the knowledge argument could use to explain how Mary might acquire such beliefs.

5 The Content of Mary's Belief

5.1 Introduction

We are investigating whether Mary acquires new propositional knowledge about types of colour experiences when she sees coloured objects. Establishing this claim requires that we first clarify the content of the belief that is involved in this knowledge. The previous chapter considered how Mary might come to believe that colour experiences have *qualia*. It emerged that Mary cannot acquire this belief by being directly aware of her colour experiences. Neither perception nor introspection can provide this direct awareness.

This chapter suggests an alternative explanation of how Mary might acquire beliefs and knowledge that colour experiences have *qualia*. Thus, the upholder of the knowledge argument might endorse this account. In particular, such an account will be elaborated by using a model of introspective knowledge offered by Fred Dretske.¹

Section 5.2 illustrates Dretske's account of introspection. Section 5.3, considers the central claim of this proposal. Namely, the capacity to have introspective beliefs about types of colour experience requires having certain perceptual beliefs and certain *connecting beliefs*. Section 5.4 shows that Dretske's account of introspection is an instance of a more general doctrine. Specifically, this account is *independent of* Dretske claims concerning the nature of conscious experience and the content of introspective beliefs. Finally, section 5.5 shows how adopting this general account of introspection illuminates our discussion of Mary's case.

¹ Although there are differences in the details, the general lines of this account of introspection are shared by Evans 1982 pp. 224-235, and Shoemaker 1996.

5.2 An Account of Introspection

A central question remains to be answered. We have to determine whether the upholder of the knowledge argument can explain how, by seeing coloured objects, Mary comes to believe that colour experiences have *qualia*.

This section shows that an account of introspection offered by Fred Dretske might suggest an answer. First, I outline the requirements that an account of how Mary acquires her supposed new beliefs about experience should satisfy. Then, I consider Dretske's account of introspection. Finally, I will show that this account satisfies these requirements. However, I will show that using this account for the evaluation of Mary's case faces two problems.

Explaining how Mary acquires her supposed new beliefs should be consistent with certain requirements concerning these beliefs. In the previous chapter, we saw two of these conditions. First, these beliefs are acquired by seeing coloured objects. Second, the transition, from seeing coloured objects to having these beliefs, cannot involve direct awareness of colour experiences. A third requirement should now be added.

Mary's supposed new beliefs concern the occurrence of properties of colour experiences that supposedly figure in our ordinary categorisation of these mental states. The reasons for this are as follows. First, Mary comes to know something about the type of colour experience she has when seeing an object of a certain colour. For example, she learns that the experience of a red object has a certain *quale*. In fact, we can exclude that she comes to know a characteristic common to all colour experiences; otherwise, she might know about it by seeing white and black objects before her release. Second, the knowledge argument seems to appeal to our ordinary understanding of what we know about colour experiences when we have them. Thus, it appears that within the knowledge argument resides an idea of *qualia* as properties of colour experiences that can determine a categorisation of

these mental states. Let us consider an account of our knowledge of colour experiences that might satisfy these requirements.

Fred Dretske's view on introspective knowledge of colour experiences is condensed in the following passage:²

What one comes to know by introspection are, to be sure, facts about one's mental life - thus (on a representational theory) representational facts. These facts are facts, if you will, about internal representations. The objects and facts one perceives to learn those facts, however, are seldom internal and never mental... One becomes aware of representational facts by an awareness of physical objects. (Dretske 1995: 40)

Let us illustrate the main assumptions of Dretske's account.

According to representationalism, the character of an experience is completely specified by its content - the ways in which the world appears to be to the subject in virtue of having that experience.³ In particular, Dretske thinks that experiences have the indicating function of providing information about the properties of external physical objects.⁴ By endorsing a form of externalism, he thinks that the representational content of veridical experiences is constituted, in part, by factors that are external to the subject. In addition, he maintains that the indicating function of sensory systems can be accounted for naturalistically. In fact, this function is acquired in virtue of evolutionary history of the organisms.⁵ In this account, our experiences provide awareness of physical objects and their properties in virtue of representing them.

² One version of the account is offered in Dretske 1995, pp. 41-44. A slightly different version has been given in Dretske 1999.

³ This doctrine is endorsed by Harman 1990, Tye 1995, Carruthers 2000.

⁴ Dretske 1995, p. 2.

⁵ Dretske 1995, p. 15.

Another assumption central to Dretske's account is that introspective beliefs that figure in introspective knowledge of experiences are *metarepresentations*. These metarepresentations, then, are beliefs that represent experiences *as* representations. Using an analogy, Dretske illustrates what distinguishes metarepresentations from other representations of representations. A photograph is a pictorial representation. As such, we can have beliefs about the weight or the geometric shape of the photograph. Although these beliefs are representations of features of an object that is itself a pictorial representation, they are not metarepresentations. Conversely, if we think about the object as a photograph of something, then we are representing it *as* a representation.⁶ For example, if I believe that the piece of paper in front of me is a photograph of a certain person, then I am regarding the object as a representation of that person. Analogously, the content of introspective belief about an experience is that such an experience is a representation of certain features of the external world.

Another assumption in Dretske's account is that introspection is a case of *displaced perception*.⁷ In certain cases, we can come to know the fact that a certain object *k* has the feature *F*, what Dretske calls a *displaced* or *target* fact, by being aware that an *intermediary* object *g* has a feature *G*. For example, we come to know that we have a certain weight by perceiving that the scale's pointer points to a certain numeral. Similarly, I come to know that the petrol tank in my car is empty by observing that the fuel gauge is pointing to "Empty". In particular, displaced perception is an inferential form of knowledge that requires connecting beliefs that relate what we are directly aware of to the displaced fact. For instance, I come to know that I have a certain weight by perceiving the scale's pointer. This requires

⁶ Dretske 1995, pp. 43-44.

⁷ Dretske 1995, pp. 41-44.

that I have beliefs about the scale's function and how the pointer's position indicates my weight.

According to Dretske, our introspective knowledge about the type of colour experiences that we have is based on beliefs about the coloured objects that we see. For instance, someone who sees a red object is in the position to believe that she is having colour experience of red. Clearly, this account is grounded on an explanation of the connection between (i) beliefs that certain objects look a certain colour to a subject and (ii) introspective beliefs that concern the type of colour experience that the subject is having.

Dretske's account offers useful insights into the type of capacities required by the subject before being able to hold beliefs about colour experiences in virtue of seeing coloured objects. A subject's ability to hold beliefs about the type of colour experiences that she is having requires her to further possess the ability to have beliefs concerning the colours of the objects she sees. Let us now consider whether this account satisfies the main requirements we have delineated at the beginning of this section.

The *displaced perception model (DPM)* of introspection explains Mary's ability to have beliefs about her colour experience in virtue of seeing a coloured object, without assuming that she is directly aware of her experiences and their properties. According to *DPM*, introspective beliefs about experiences function in the same way as the beliefs about displaced facts. We are directly aware of the facts and properties represented by the experiences; and on the basis of this awareness and appropriate connecting beliefs, we acquire introspective beliefs about our experiences. Let us now consider whether *DPM* offers an account of what is required by our ability to categorising colour experiences introspectively.

The *DPM* account clarifies introspective categorisation of colour experiences by revealing that two types of property are involved in one's ability to have introspective beliefs. The properties of the first type are required by the analysis that

DPM provides of the content of introspective beliefs concerning experiences. As Dretske suggests, introspective knowledge involves metarepresentations about our mental experiences. This means that we think about our experiences as states that represent the world in a certain way. Let us consider the introspective belief that:

(1) I am having an experience that a certain object is red.

The content of this belief is analysed as self-ascribing a state that has the property of being the representation, or awareness, that something is red.⁸ Therefore, we ascribe to an experience the property of making us aware of something.

According to *DPM*, there are also properties of another type that play an important role in the introspective categorisation of colour experiences. Our experiences make us aware of these properties. For example in seeing red, we are aware of redness. As we have seen, by being directly aware of properties of redness we categorise our experience as an experience of red.

These two types of property appear to satisfy different conditions usually associated with the notion of *qualia*. According to a common use, *qualia* are properties of experiences. Therefore, we might assume that properties such as “being a representation, or awareness, of red” are *qualia*. On the other hand, some assume that *qualia* are features of which we are directly aware and as such ground our introspective classification of experiences. In this case, it seems that the properties of which our colour experiences makes us aware should be called *qualia*. Choosing between these two interpretations of *qualia* is a question of terminology. What it is important is that the *DPM* reveals that there are two types of property involved in our introspective knowledge and classification of colour experiences.

The *DPM* appears to elucidate some central features of the transition occurring between seeing a coloured object and having a belief about the *qualia* of the

⁸ The idea that colour experiences are states of p-awareness of properties can be found in Dretske 1999.

experience. Therefore, the upholder of the knowledge argument might endorse the following account:

(3) Mary comes to believe *that* experience *e* has *quale* *Q* because, in virtue of seeing coloured object *r*, she believes that the object *r* looks a certain colour to her and she has some appropriate connecting belief.

However, the viability of this suggestion depends on answering two important questions.

The upholder of the knowledge argument needs to show that Mary's supposed new beliefs about colour experiences amount to knowledge. In determining how this can be done, we have to consider the *DPM* in more detail. We have to investigate how the connecting beliefs required by *DPM* warrants beliefs about the *qualia* of experience.

A second problem might derive from the *DPM's* connection to representationalism. Dretske's view on introspection is based upon three assumptions. First, colour experiences are typed exhaustively by their representational content. Second, given the meta-representational nature of the content of introspective beliefs about colour experience, introspection provides a categorisation of colour experiences in terms of their representational content.⁹ Third, the representational properties of experience can be accounted for in naturalistic terms. Clearly, the upholder of the knowledge argument cannot endorse this latter claim. Thus, we might think that he might endorse the idea that the representational features of experiences are non-physical properties. However, even if such a doctrine is tenable, there might still be a problem stemming from the other assumptions.

⁹ The upholder of the knowledge argument might even assume that experiences are completely individuated by their representational content. Moreover, he can also concede that introspective beliefs are metarepresentations. Clearly, what he has to reject is Dretske's naturalistic account of this content.

As Ned Block has pointed out, whether conscious experiences have features that go beyond their representational or intentional content is at the core of “The greatest chasm in the philosophy of mind -- maybe even all of philosophy --”.¹⁰ Moreover, representationalists have different views on how the notion of representational content should be spelled out. Thus, it might appear that assuming *DPM* in the clarification of how Mary acquires her new beliefs about colour experiences might require the demanding defence of representationalism.

To sum up, it seems that Fred Dretske's account of introspection might offer some insight into how Mary acquires beliefs about colour experiences. Nevertheless, before assessing whether this is the case, two main issues remain to be investigated. First, we have to determine in detail how these beliefs amount to knowledge. Second, we have to consider whether accepting *DPM* involves endorsing Dretske's representationalism. These questions will be addressed respectively in the following two sections.

5.3 Connecting Beliefs

The *DPM* appears to offer a promising account of the way in which, upon her release, Mary acquires her beliefs about colour experiences. The central assumption in *DPM* is the existence of certain connecting beliefs. These beliefs are supposed to ground the transition between beliefs concerning the coloured objects we see and the types of colour experiences we can self-ascribe. Moreover, these transitions are supposed to deliver knowledge about colour experiences.

This section aims to investigate the nature and role of these connecting beliefs. Specifically, I will suggest that these beliefs concern the relation of an ordinary notion of a certain type of colour experience with that of an object looking a certain colour.

¹⁰ Block 1996.

According to Dretske, in knowing a fact by displaced perception we are not directly aware of that fact, instead we come to know about it in virtue of an inference that involves a *connecting belief*. In consulting a scale, for instance, we do not perceive ourselves having a certain weight, but we perform an inference of the following type.¹¹ For example, let us assume that we know that:

- (1) The scale points to a certain numeral n .
- (2) The scale would not point to numeral n , unless we weighed n kilos.

From these premises, we conclude that:

- (3) We weigh n kilos.

Whilst assumption (1) specifies what we are directly aware of, (2) gives the content of a connecting belief, and the conclusion (3) is what we come to know about the displaced fact. In order to see that we have a certain weight by seeing the scale, we need to believe that the position of the pointers is a “sign” of our weight. In such a case we need to know that the position of the pointer is causally determined by our weight and thus that a principle like (2) holds.¹² Let us now consider how this model might be applied to introspective beliefs.

Dretske does not present explicitly the inference involved in introspective knowledge of colour experiences. Thus, he does not give an example of the connecting belief involved in this inference. However, some of his passages suggest how this principle might be. In fact, he claims that:

One comes to know (the fact) that one is experiencing blue by experiencing, not the experience of blue, but some displaced object. ..., this displaced object is (typically) the object the experience of blue is an experience of – i.e. the blue object one sees. (Dretske 1995: 44)

¹¹ Dretske 1995, p. 42.

¹² Dretske 1995, p. 92

Moreover, he states that:

If you “see” *k* as blue and infer from this “fact” – the “fact” that *k* is blue – that you are representing *k* as blue, you cannot go wrong. As long as the inference is from what you “see” *k* to be (whether this is veridical or not) the conclusion must be true: blue must be the way you are representing *k*. (Dretske 1995: 61)

Let us outline the inference that might be involved in the introspection of a colour experience.

Some authors have offered an interpretation of Dretske’s account.¹³ They assume that the *DPM* requires an inference from what we are aware of to the belief that we are having a certain type of colour experience. Specifically, they assume that the starting point of this inference is the content of a perceptual belief. This belief is acquired in virtue of having the experience. Let us assume that I see a red object. According to this interpretation, the first premise of the introspective inference is the content of the belief that:

(1) This object is red.

Now, following the analysis of the displaced perception of our weight, I should have the connecting belief that:

(2) This object would not be red unless I were having the experience that this object is red.

Given these two premises, I can infer:

(3) I am having the experience that the object is red.

¹³ Lycan 2003 and Aydede 2000.

The plausibility of this account, of course, depends on the assumption that the content of the connecting belief is expressed by (2). However, this account appears to have some problems.¹⁴

If we endorse an *objectivist* account of colour properties, the connecting belief (2) is clearly false. According to the objectivist, colours are physical properties of the external objects (or their surfaces) whose occurrence and existence is independent of our sensory responses. In this case, an object can be red even if no one perceives it. Therefore, principle (2) turns out to be false and it cannot support the inferential introspective knowledge of colour experiences.

The connecting belief (2) might appear to be more plausible if we endorse a subjectivist account of colour properties. According to this metaphysical view on colour properties, an object has a certain colour property only if it elicits a certain response or experience in a subject. In this case, the statement “this object is red” is true only if the object determines a certain colour experience in a subject. However, it does not seem that a subject *S*, who believes that something is red, is justified in concluding that she is having a certain colour experience. In a subjectivist account of colour, a certain object has a certain colour when it determines a type of response defined over a range of subjects. Thus, an object might be red in virtue of the response of some subject different from *S*. Therefore, *S* cannot be justified in thinking that: “This object would not be red unless I were (probably) having the experience that this object is red”.

If the present interpretation of Dretske's account is correct, his view is not tenable. Clearly, the correctness of this interpretation might be questioned. However, we can put this issue aside.¹⁵ What is relevant here is to see whether there

¹⁴ Criticisms of this account of introspective connecting beliefs are advanced both in Lycan 2003 and Aydede 2000.

¹⁵ It is important to notice that in the subsequent paper Dretske 1999, although Dretske defends the idea that we have indirect knowledge of colour experiences, he makes no mention of connecting principles.

is another way to spell out the general intuition involved in the *DPM* of introspection.

There is an alternative account of how we form the introspective belief that we have a colour experience of a certain type by seeing a coloured object. Let us assume that subject *S* sees a car. Given what she sees, *S* might come to believe:

(1) This car looks blue to me.

In the proposed account the connecting belief would be:

(2) This car would not look blue to me unless I were having a colour experience of a certain type.

S can reach the introspective belief:

(3) I am having a colour experience of a certain type.

Let us clarify this account, starting with the belief that (1).

Perceptual beliefs of the form “this object is *P*”, where *P* stands for a colour predicate, are not the only type of beliefs we might endorse in virtue of seeing a coloured object. Another class of beliefs we can have when we see coloured objects have the form “this object *looks P* to me”. An important difference between these two types of belief is that we might be ready to endorse the latter without endorsing the former. Let us assume that, given all my past experiences, I believe that my car is red. If I see the car under a sodium-arc streetlight, that produces a distinctive yellow light, I might be in the position to believe that the car is blue. However, given that I have evidence that the car is red, I will not be ready to believe that the car is blue. In this case, I will maintain that the car looks blue. Let us consider now the connecting principle (2).

It seems that beliefs concerning the colours object look to have, contrary to those concerning the colour they do have, are connected in some systematic way with the notion of colour experience. It makes sense, then, to explain to a subject

that a car looks blue to her because she is having an experience of it under certain conditions. The connecting beliefs that, according to *DPM*, are required in indirect introspective knowledge of colour experiences might concern these connections.

In particular, it might be suggested that principle (2) is grounded in what is required for the possession of the concept of colour experience. The knowledge of certain inferential patterns, where the concept of experience might figure, is a necessary condition for possessing the concept.¹⁶ Possessing the concept of colour experience requires knowing that colour experiences figures in the explanations of why, under certain circumstances, things look to have certain colours to one. In particular, this would require that something look a certain colour to someone because she has a colour experience of a certain type. Therefore, let us assume that someone does not know how to use the notion of colour experience to explain why a certain object looks to be a certain colour to him. Under such circumstances, this subject can be said to lack the notion of colour experience. For example, someone cannot be said to possess the concept of the experience of red, if he does not know that a certain object would not look red to him unless he had that experience. Thus, the connecting belief, which figures in a subject's introspective classification of a type of colour experience, derives from his mastery of the concepts of that type of colour experience.

It might be objected that this version of *DPM* provides a circular account of the inference involved in introspection. In fact, it might be claimed that someone cannot believe that an object looks blue unless he believes that he is undergoing an experience of blue, thereby making the previous inference circular. However, it seems that this challenge can be met.

¹⁶ The relation between possessing concepts and being able to find compelling certain inferential schema has been investigated by Peacocke 1992 see also Millar 1991a.

A subject comes to believe that an object looks a certain colour because she realises that, despite what she sees, she is not in the position to judge that the object has that colour. However, in order to realise this, she does not require having beliefs about her own colour experiences. She might withdraw the assent to her belief that the object is a certain colour given that this belief contradicts other beliefs she has already acquired about the colour of the object. In the case where she is seeing the red car under the yellow light, she might not believe that “the car is blue” because from past experiences she acquired the belief that the car is red. It seems plausible that she can withdraw her assent to the belief “the car is blue” and endorse the weaker “the car looks blue” without knowing anything about her experiences. She might just adduce as the reason for having the belief that “the car looks blue” some puzzling change in the surface of the car. Thus, knowing that her experiences are involved in determining the colour the object looks to have would be informative for her.

Having illustrated how *DPM* might explain the introspective knowledge of experience, we have now to consider whether this account requires endorsing a representational theory of experience.

5.4 Representationalism and Introspection

In this section I will illustrate that the general intuition involved in the *DPM* does not require endorsing a representationalist account of introspection. In fact, I will argue that this account is consistent with a non-representationalist account of experience.

Dretske's account of introspection involves two main assumptions. First, colour experiences are typed by their representational content. Thus, two colour experiences are of the same type when they have the same representational content. Second, our introspective beliefs about experiences take into account this typology of experiences. According to Dretske, in self-ascribing a type of colour experience

we have to think about it as a type of representational state. Let us assume, for instance, that Mary sees a red object. On this account, Mary should have a belief about her colour experience expressible as:

- (1) I am having a colour experience that represents that something is red.¹⁷

The *quale* *Q* is the property of “being the representation that an object looks red”. In fact, according to Dretske, experiences are typed by their representational content.

However, representationalism is a debated position. Thus, it might appear that assuming *DPM* in the clarification of how Mary acquires her new beliefs about colour experiences might require the demanding defence of representationalism. If we renounce representationalism, then introspective beliefs about experiences cannot individuate experience as representations nor type them in terms of their representational content specified by basic perceptual beliefs.

Thus, we can endorse the idea that introspective beliefs derive from an inferential procedure starting from certain beliefs that do not concern the experiences themselves but their objects. Having this grasp of the “skeleton” of *DPM*, let us see how the “flesh” of a non-representationalist account of experience can be added.

Christopher Peacocke has formulated influential arguments for the conclusion that experiences, besides having their representational features, have properties he calls *sensational*.¹⁸ According to Peacocke, the main tenet of externalist representationalism is what he calls the *Adequacy Thesis* (AT). This principle states that a complete characterisation of an experience can be given by prefixing an operator “It visually appears to the subject that ...” to some complex condition

¹⁷ In a subsequent paper, Dretske claims that the type of colour experience can be specified as the experience that has the property of “being awareness of the property of being red”, Dretske 1999, pp. 112-114.

¹⁸ Peacocke 1983.

concerning physical objects stated propositionally. If, for instance, subject *S* has an experience of red, the content of the experience is completely characterised by “It visually appears to *S* that there is a red object”. Let us examine one of the examples he advances to argue against AT.

Peacocke asks us to consider two trees of the same size that are at a distance of one and two hundred metres respectively from an observer. The observer sees the two trees *as* having the same size. According to the adequacy thesis, the representational content of the observer's perception is that the two trees coincide in size. However, it is also true that there is a sense in which we can say that the closer tree occupies a larger part of the observer's visual field than the one occupied by the more distant tree and this difference is reflected in the phenomenology of the experience. According to Peacocke, this is a phenomenological difference determined by a sensational feature of the experience.¹⁹

Peacocke maintains that sensational properties have a role in the classification of colour experiences. In his view, two experiences having the same (or similar) sensational properties are of the same type. Therefore, a sensational feature determines the type to which an experience belongs. In particular, he argues that sensational properties and representational features determine two different categorisations of experiences. He formulates arguments for the existence of experiences that match in representational content and differ in sensational properties and *vice versa*. This shows clearly how his account departs from the representationalist account of experience. However, the role he assigns to sensational properties in the categorisation of experiences has another important implication for our discussion of Mary's belief that experiences have *qualia*.

¹⁹ This conclusion is accepted here for the sake of the argument. A reply to it can be found in Carruthers 2000, p. 117.

It has already been stressed that one of the roles assigned to *qualia* is that of being properties of experience that determine our classification of colour experience. It is clear that Peacocke's sensational properties satisfy this feature of *qualia*. Thus, we might think that Mary's supposed new belief concerns the fact that the experience she is having has a certain sensational property.

Determining whether *DPM* can account for the formation of introspective beliefs that our experiences have sensational properties will answer this question. If sensational properties of experience exist, we should be able to think introspectively about them. Our task is now to see whether such introspective beliefs and the resulting categorisation of experiences might result from an inferential procedure as suggested by *DPM*. Determining how introspective beliefs about sensational properties are formed requires investigating how we can acquire concepts referring to sensational properties.

Peacocke has provided an account of how to conceptualise sensational properties. He thinks that:

The sensational properties of an experience, like its representational properties, have reliable and publicly identifiable causes. (Peacocke 1983: 305)

In particular, Peacocke suggests a way of individuating the sensational features of colour experiences that renders perspicuous the relation between thinking about these features and our beliefs about their causes. Using primed predicates to indicate the sensational features of experiences, he suggests that we can fix the reference of red' to the relative sensory property in term of this description: "Red' is the property of the visual field in which a red thing is presented in normal circumstances". Let us see how this way of fixing the reference to sensational property can be used to provide a *DPM* account of our introspective knowledge of the type of experiences we have.

Given the sensationalist account of colour experiences, certain introspective beliefs about the type of experiences we are having might be analysed as “I am having an experience with sensational property *P*” where *P* stands for a sensational property like *red*, *green* etc. Thus, in order to have an introspective belief we need to possess the concept of an experience of a certain type. In turn, if the sensationalist account of experience is correct, we need to possess a way of thinking about sensational properties.

We have seen that, according to Peacocke, we can think about sensational properties in terms of certain reference fixing phrases concerning objects of the relevant type that cause certain experiences. Such relations of dependence in the possession of these concepts suggest a way of spelling out the introspective knowledge of experiences in terms of the *DPM*. Let us consider, for instance, the sensational property *being elliptical*. This is the sensational property normally caused by an object presenting an elliptical aspect to the observer. We learn to ascribe such a property to our experience by noticing elliptical aspects. For example, let us assume that I believe that the experience I have, as I look at my coffee mug, has the sensational property of *being elliptical*. Having this belief can only be based on the fact that I notice that the mug rim presents an elliptical aspect from my point of view. Then I assume that its doing so affects the character of my experience. Similarly, we can apply this account to the sensational properties of colour experiences. Let us assume that a subject comes to believe that his experience has a certain sensational property *red*. It seems plausible that the subject forms this belief by noticing that an object looks red to him in the given observational conditions.

To sum up, we can regard introspection as indirect awareness of experiences. Specifically, we can formulate this idea without having to choose between representationalism and sensationalism. We have now to consider whether this

account of introspection is specific enough to describe how Mary comes to know that experiences have *qualia*.

5.5 The Content of Mary's Belief

The *DPM* explains how we ordinarily form beliefs about colour experiences when we see coloured objects. Moreover, this account clarifies how these beliefs amount to knowledge. The *DPM* does not require endorsing a particular view on the nature of colour experiences. This doctrine is based on a plausible claim concerning our introspective capacities. According to this general principle, subjects can categorise colour experiences introspectively only if they have two capacities. Firstly, they have to form beliefs about colours that objects look to have to them. Secondly, they should possess a typology of colour experiences that relate a type of colour experience to the colour (shade) that they are seeing. This section considers how endorsing *DPM* affects our evaluation of the knowledge argument.

According to the upholder of the knowledge argument, upon her release, Mary discovers what it is like to have a type of colour experience. This discovery involves propositional knowledge concerning the type of colour experience that she has when she is seeing a coloured object. This means that when Mary sees a coloured object she discovers that the scientific description of the type of colour experiences she is having is incomplete. There is a property left out by her scientific knowledge. As some authors illustrate this point, there should be a discovery that Mary can formulate as follows: "Aha, this is what it is like to see red!".

To make such a discovery, Mary should have two capacities. First, she has to be able to determine which type of colour experience she is having. This means that she has to recognise that her current colour experience is of a type characterised by scientific knowledge. Second, she has to discover the characteristic of this type of colour experience that she could not know before her release. It appears that the *DPM* explains how Mary can have these capacities.

We can account for how Mary recognises that her current colour experience belongs to a certain type that is described in scientific terms. Before her release, Mary can describe what it is to look red.²⁰ Firstly, Mary possesses a schematic notion of looking a certain colour. By seeing black and white objects, she is able to understand that objects can look to her to have certain colours. Moreover, she can describe red in scientific terms. Given our account of Mary's scientific knowledge, she can describe red as a certain position *P* in the complete colour space. Thus, she understands red as the feature that satisfies certain relations of similarity to the other colours. Alternatively, she can characterise red as the colour of a certain paradigm stimulus seen in certain conditions. Moreover, she can describe red as the colour that objects look to have to people when they are in a certain brain state. Whichever scientific description Mary uses to characterise red, let us indicate it with *Red_s*. Accordingly, let us use the expression *Red_s-type experience* to indicate a description that Mary can use before her release to describe a type of colour experience. Namely, this is the description: "The type of colour experience someone has when something looks *Red_s* to him".

The *DPM* account suggests a plausible explanation of how Mary acquires introspective beliefs concerning the type of colour experience she is having. Let us assume that Mary sees a red rose. On this account, Mary should be able to draw the following inference:

- (1) The rose looks *Red_s* to me.

- (2) The rose would not look *Red_s* to me unless I were having a *Red_s-type* experience.

Therefore:

²⁰ Of course, here I am not assuming that such a description is satisfactory.

(3) I am having a *Red_s-type* experience.

It seems plausible that Mary can know (1) and (2). Before her release, Mary can know that roses look *Red_s* to normal individuals in certain visual conditions.²¹ When she sees the rose in the appropriate condition, she will recognise that it is a rose. Thus, she can conclude that the rose looks red to her. Moreover, nothing bars Mary from knowing principles such as (2). Before her release, she can possess the concept of visual colour experience. By seeing black objects she can infer that she is having a *black-type* colour experience. This means that she can form the belief that she is having the type of colour experience people have when things look black to them. Thus, she can master both a “schematic” notion of colour experience and that of looking a certain colour to her.

Thus, we can understand how Mary can self-ascribe a type of colour experience as described by her scientific knowledge. Clearly, the knowledge argument is not meant to raise a problem concerning this self-ascription. Instead, this argument is intended to show that, by seeing a coloured object, Mary learns something about this type of colour experience that escapes her scientific knowledge. Let us, then, consider this issue.

The *DPM* clarifies how Mary forms a belief that might plausibly amount to the knowledge of what it is like to have a certain type of colour experience. By seeing the rose, Mary comes to believe something else beside (1). In fact, she can also acquire the belief that:

(4) The rose looks *this* colour.

Moreover, given the belief that (1), Mary can form the belief that:

(5) Something looks *Red_s* when it looks *this* colour.

²¹ A suggestion of this type is formulated in Perry 2001, p. 96.

Finally, she can form a belief about the *Red_s-type* experience that she is having. Namely, she can believe that:

(5) A *Red_s-type* experience is the type of colour experience one has when something looks *this* colour to her.

Thus, it might be suggested that (5) is a plausible candidate for the type of belief that should figure in the knowledge of *what it is like to have a certain type of colour experience*. Moreover, this account suggests a characterisation of the *quale* that Mary comes to know. On this account of introspective knowledge, Mary comes to know that her colour experience has the property of “being the type of colour experience one has when something looks *this* colour to her”.

The upholder of the knowledge argument might object to such a revision of the concept of *quale*. In fact, our characterisation introduces a description of *qualia* that cannot figure in the original setting of the knowledge argument. There are two lines of reply to this objection. The first reply is that the suggested interpretation derives from the account of introspection that was defended in the previous sections.²² Therefore, the burden of proof is on the promoter of this criticism. He has to show that this account of introspective knowledge is implausible.

The second reply to the supporter of the knowledge argument is that our revision of the notion of *quale* does not undermine the knowledge argument. The intuition that Mary learns something about her experience is preserved. On our account, Mary learns something new about the type of colour experience she is having, only if she learns something about the colour *Red_s*. In fact, the notion of the *quale* of the experience under discussion here is formulated with reference to whatever Mary learns once she has the demonstrative belief (4). Moreover, there is a plausible intuition that, in knowing (5), Mary learns something. First, as we saw in

²² See sections 5.2 and 5.3.

chapter 3, although Mary knows the precise position of the colour in a system of relations of similarity she cannot recognise that the rose looks *Red_s* to her just by seeing the rose.²³ Clearly, if we want to trick her and we show her a blue rose she would believe that this rose looks *Red_s* to her. Similarly, she will not be able to determine amongst different randomly coloured patches which one is *Red_s*.

Thus, the upholder of the knowledge argument might argue that these new recognitional and discriminatory capacities depends on the instantiation of non-physical properties. For instance, he might maintain that Mary's discovery that objects that look *Red_s* have some properties that she did not know before her release. Thus, she is able to recognise or discriminate by sight *Red_s* things in virtue of her awareness of these properties. The upholder of the knowledge argument might argue that in seeing the rose Mary comes to know something new.

To sum up, the central problem we have to face in evaluating the plausibility of the central step in the knowledge argument is now clear. We have to establish whether, by seeing a coloured object, Mary acquires a new belief concerning the colour that the object looks to have to her. In fact, this is a central requirement for Mary acquiring a new true belief about the type of colour experience that she is having. Moreover, only if the upholder of the knowledge argument can prove this claim, he can then proceed in arguing for the conclusion that there are facts that Mary's scientific knowledge cannot accommodate.

²³ I avoid considering the other types of scientific description that Mary might use to characterise *Red_s*. It seems that similar considerations apply in the case that Mary uses them. If for example she describes something that looks *Red_s* as looking the colour of a certain paradigm stimulus, she will not be able to recognise that the rose looks *Red_s*. Similarly, this will happen for any physical specification of the nature of the stimulus or of the brain states she might use to articulate the notion *Red_s*.

5.6 Conclusion

When considering the content of Mary's new knowledge, Jackson and other commentators claim that she comes to know about *qualia* as features of her experiences. However, the thought experiment alone cannot support this conclusion. Conversely, I have argued that Mary is directly aware of the objects of her experiences and their features. Moreover, we have an account of how she can think about her colour experiences when she sees coloured objects. This model does not affect the idea that upon her release Mary might learn about new facts and properties.

6 Resisting the Ontological Conclusion

6.1 Introduction

We have to establish whether a modified version of the knowledge argument raises a difficulty for the hypothesis of modest reductionism. The knowledge argument is based on the claim that knowing what it is like to have a colour experience escapes Mary's scientific knowledge.

The previous chapters elucidated the two kinds of knowledge of colour experiences that are involved in this version of the knowledge argument. The third chapter presented an account of Mary's scientific knowledge. This knowledge relies on a descriptive apparatus that we can intelligibly grasp as a generalisation of models employed in contemporary science.

The fourth and the fifth chapters investigated the knowledge that, according to the supporter of the knowledge argument, Mary acquires upon her release. Specifically, I offered an account of the belief that figures in Mary's knowledge of what it is like to have a colour experience.

This chapter considers the central inferential step involved in the knowledge argument. This step can be expressed as an instance of *modus ponens*. Namely: if, by seeing coloured objects, Mary acquires new propositional knowledge about colour experiences, then there are facts concerning these mental states that her scientific knowledge cannot accommodate. Moreover, Mary acquires new propositional knowledge about colour experiences.

Firstly, I will evaluate whether the latter claim is true. This claim requires that Mary's knowledge of what it is like to have a colour experience is based on *new* true beliefs. This means that Mary could not have these beliefs about colour experiences before her release. In the previous chapter, we established that Mary

acquires *new* beliefs about her colour experiences, *only if* she acquires *new* beliefs about the colours that objects look to have. Thus, we have to consider whether the antecedent of this conditional is true.

Section 6.2 begins this investigation by discussing the *ability reply* to the knowledge argument. According to this criticism, by having colour experiences Mary simply acquires a set of abilities or forms of *knowing how*. However, she does not acquire beliefs about colour experiences. Therefore, she does not acquire new beliefs. Section 6.3 shows that this reply might be effective against the idea that Mary gains new propositional knowledge about colours. In section 6.4, I argue that an evaluation of the ability reply requires possessing a criterion for the individuation of beliefs. Then, I show that the upholder of the knowledge argument can evade the ability reply by endorsing a certain criterion for the individuation of beliefs. Consequently, we can concede that Mary acquires true beliefs about colour experiences that she lacked before her release.

The remainder of the chapter considers whether making such a concession requires accepting the conclusion of the knowledge argument. This means establishing whether, if we assume that Mary acquires new propositional knowledge, we must accept the existence of facts concerning her colour experiences that science cannot account for.

Section 6.5 illustrates an objection to this conclusion that I call the *two ways of thinking reply*. The supporters of this reply assume that the premises of the knowledge argument are consistent with the possibility that Mary acquires, upon her release, a new way of knowing facts about colour experiences she already knew before her release.¹ This means that although all the premises of the argument might be true the conclusion that there are non-physical facts does not follow. Therefore,

¹ Recent versions of this strategy against the knowledge argument are pursued by Carruthers 2000, Horgan 1984, Lycan 1995, Rey 1992, Tye 2000, Papineau 2002, Sturgeon 2000, Perry 2001, Loar 1990.

the knowledge argument is not valid. Section 6.6 shows that such a reply requires the existence of certain concepts, usually called *phenomenal concepts*, concerning colour experiences.

6.2 The Ability Reply

The knowledge argument requires that Mary acquires new true beliefs and knowledge about colours by seeing coloured objects. Only if this condition is satisfied, does Mary acquire new knowledge about colour experiences. Moreover, only if this latter claim is true, the upholder of the knowledge argument can draw her conclusion. Namely, there are facts concerning colour experiences that Mary's scientific knowledge cannot account for.

This section illustrates the *ability reply* to the knowledge argument proposed by Laurence Nemirow and David Lewis.² They argue that Mary does not acquire new beliefs about her colour experiences once she sees coloured objects. Instead, she gains certain abilities.

The ability reply is an attempt to block the crucial step in Jackson's knowledge argument. This criticism aims to show that if Mary learns something new, the existence of non-physical facts does not follow. The strategy involves showing that knowing what it is like to have an experience is not knowledge that requires that certain facts hold. Two main premises are involved in arguing for this conclusion. The first premise is that, by seeing colours, Mary acquires only a set of abilities or a form of *knowing how*. Let us call it the *ability hypothesis*. The second premise is that propositional knowledge and knowing how are different kinds of knowledge. Knowing how does not involve acquiring beliefs whose truth requires that certain facts hold. Thus, there is no reason to conclude that Mary acquires beliefs about

² Nemirow 1980, Nemirow 1990, Lewis 1983, Lewis 1990. Mellor 1993 provides a similar account of knowing what it is like to have an experience, but without defending physicalism.

facts that she could not know before her release. Let us consider these premises in more detail.

The *ability hypothesis* states that Mary, by seeing colours, simply acquires certain abilities. Nemirow, for instance, advances the ability hypothesis as the following equation:

Knowing what it is like is the same as knowing how to imagine having the experience. (Nemirow 1990: 495)

He maintains that knowing what it is like to have a colour experience is correlated with imagining the sight of a colour. According to Nemirow, Harry cannot honestly affirm that he knows what seeing chartreuse is like if he cannot imagine the sight of chartreuse. Similarly, Nemirow claims that: “It would be nonsense for Harry to insist that he can easily visualise chartreuse, but does not know what seeing it is like”.³ Nemirow suggests that this correspondence should be taken as the manifestation of an identity. He maintains that the ability hypothesis should be endorsed for its explanatory powers. In particular, he thinks that this equation explains the central intuitions involved in the knowledge argument in terms compatible with physicalism.⁴ However, before investigating whether this is the case, let us consider another formulation of the ability hypothesis.

Lewis formulates the ability hypothesis with a more extensive list of abilities:

... knowing what it is like is the possession of abilities: abilities to recognize, abilities to imagine, abilities to predict one's behaviour by imaginative experiments. (Lewis 1983: 131)

Besides the ability to imagine suggested by Nemirow, Lewis adds both the ability to recognise and to make certain predictions. According to Lewis, once we have an

³ Nemirow 1990, p. 493.

⁴ Nemirow 1990.

experience we acquire the ability to recognise it again. If we taste a certain food, we acquire the ability to recognise whether on another occasion something tastes the same way and thus that we have the same experience. Moreover, having a certain experience enables us to predict how we would react if we were to have a similar experience. For example, if we taste a disgusting food, this experience enables us to predict that we will avoid such a gustatory experience in the future.⁵

The second premise of the ability reply is the endorsement of the distinction between *propositional knowledge* and *knowing how*.⁶ Propositional knowledge involves true beliefs. A belief can be regarded as subjects' attitude towards a proposition. Propositions are representations of reality that can be either true or false. Thus, a belief can be said to be true (false) when the relative proposition is true (false). Moreover, it is plausible to assume that implicit within the knowledge argument is the idea that the obtaining of facts renders propositions and beliefs true. Given this characterisation of propositional knowledge, we can appreciate the main difference between this kind of knowledge and knowing how. Possessing abilities, such as knowing how to swim, catch a ball or to play musical instruments, is a form of knowledge. However, knowing how to do something is not to believe that the world is a certain way. Thus, this knowledge does not involve propositions that can be either true or false. From this, it follows that this knowledge does not concern facts. In order to exercise the relevant ability, certain conditions in the world have to be satisfied. Nevertheless, having these abilities does not amount to knowing that certain facts occur.

The upholder of the ability reply offers a direct response to the knowledge argument. By seeing coloured objects, Mary acquires abilities and these abilities *qua* knowing how do not involve new beliefs. Without these new beliefs about

⁵ Lewis 1983, p. 131.

⁶ A classical analysis of the differences between *knowledge that* and *knowing how* is given in Ryle 1949, pp. 26-60.

colour experiences, there are no facts concerning properties that her scientific knowledge cannot accommodate. At the same time, the supporter of the ability reply can explain the intuitions exploited in the knowledge argument. First, she can explain why knowing what it is like to have an experience requires having that mental state. Seeing a colour is a condition for acquiring the ability to imagine, recognise and remember the relative colour experience. Second, she can account for the intuition that, upon her release, Mary learns something new. Before her release, Mary has knowledge that might be acquired just by being told about certain facts. However, simply being told about facts does not give her the abilities that constitute knowing what it is like to see a colour. Consider someone who is told all the facts involved in swimming. This would not give him the ability to swim; he needs practice. We have now to consider whether the ability reply is relevant for establishing whether Mary acquires a new belief about objects looking the same colour.

To sum up, the ability reply denies that Mary acquires a new propositional knowledge about colour experiences. On this view, she just acquires certain abilities. Now we should consider whether this reply might apply to our version of the knowledge argument.

6.3 A Version of the Ability Reply

This section shows how the ability reply applies to our version of the knowledge argument. According to the ability reply, when Mary sees a coloured object she acquires amongst others abilities a recognitional ability. This is the ability to recognise a type of colour experience when one has it again. I will show that if Mary acquires this ability, it can be shown that she cannot acquire a new true belief concerning the *quale* of her colour experience. Before doing this, however, we need some preliminary clarifications and assumptions.

In the previous chapter, I characterised what might be the *quale* that upon release Mary supposedly ascribes to her colour experience. Let us assume that Mary sees a red rose. In addition, we can use *Red_s* to stand for Mary's scientific description of red. Finally, *Red_s-type experience* is an abbreviation for "the type of colour experience that people have when something looks *Red_s* to them". Therefore, by seeing the rose Mary acquires (supposedly) the knowledge that the *Red_s-type experience* has a *quale*. I suggested that Mary can regard this *quale* as "the property that determine the type of colour experience that one has when something looks *this* colour to him".

There are two plausible assumptions we have to consider. First, if Mary comes to have the new true belief that a type of colour experience has a *quale*, then, in certain circumstance, she will acquire other new true beliefs. She will have new true beliefs when judging that two colour experiences are of the same type. For the supporters of the knowledge argument assume that *qualia* are properties that ground the ordinary typology of colour experiences. On this view, two colour experiences are of the same type when they have the same *quale*. The second assumption concerns the capacity to judge that two colour experiences that one is having are of the same type. This requires judging that the objects one sees look the same colour. Clearly, this follows from our *DPM* account of introspective knowledge. We are now in the position to consider how the ability reply can be used against our version of the knowledge argument.

Let us suppose that Mary, before her release, is allowed to study a coloured patch **A** without seeing its colour. In particular, Mary knows the physical properties of the light reflected by **A** and the kinds of responses subjects have to it as described by her scientific theory. Thus, she can locate the evoked response to **A** in the complete colour space. Therefore, she can know all the relations of similarity that

this response has with all the other responses to light stimuli.⁷ Moreover, she can explain in neuroscientific terms why the response evoked by seeing **A** occupies that position it does in the colour space.

In particular, Mary can predict which stimuli are *metamers* of the light stimulus relative to **A**. Metamers are physical stimuli that evoke judgements of colour matching. Metameric stimuli can differ completely in their wavelength composition.⁸ In particular, for any stimulus at a certain wavelength it is possible to determine a second stimulus of different composition that will produce a matching colour response. This second stimulus has to be composed of appropriately chosen wavelengths and intensity. Moreover, it is already possible to predict whether two light stimuli are metamers once their wavelength composition and intensity are known.⁹

Having granted that Mary knows all the relevant scientific facts about the coloured patch **A**, let us now consider two stories about Mary. In the first one, Mary is released and we show her a patch **B** whose reflected light is a *metamer* of the light reflected by **A**. However, she is not allowed to see the colour of the patch **A**. In addition, Mary is not allowed to study **B** using her scientific instruments.

In this case, it seems that Mary cannot determine, just in virtue of seeing **B** and her knowledge about **A**, whether **A** and **B** look the same colour and that they are thus *metamers*. In fact, she cannot infer that **A** and **B** look the same colour from her scientific knowledge of the properties of **A** and the mathematical procedure used to determine the *metamers* of **A**. First, she cannot conclude that the two patches look the same colour from the premise that they both look a certain colour as described

⁷ See on colour spaces section 2.3 at p. 48.

⁸ Usually, metamers are defined with respect to a certain observer under certain viewing conditions. Given that this specification does not affect my exposition, I have omitted it. On metamers see Hardin 1988, pp. 27-28.

⁹ These predictions are based on mathematical procedures involving certain models called *wavelength mixture spaces*. These models concern the response functions of the receptors in the retina. See Clark 1993, pp. 37-42.

in the colour space. In fact, just by experiencing **B**, she cannot determine the relations of similarity that **B**'s colour has with all the other colours.¹⁰ The patch might occupy all her visual field, and thus she might only know about the relations of similarity that the colour of the patch has with white and black. Second, Mary will recognise neither the physical properties of **B** nor the brain states involved in seeing it. She has never had colour experiences and she cannot know, just by experiencing **B**, how the colour she sees relates to these physical properties.¹¹ We can now consider a second story about Mary.

As in the previous case, Mary knows the features of patch **A** in terms of her scientific knowledge. But after her release, we show her patch **A** and then patch **B**. It seems that now she is in a position to recognise that **A** and **B** look the same colour. Thus, Mary can express her belief by affirming that "patch **A** and **B** look the same colour". Moreover, it seems that Mary can acquire this belief by having a colour experience. Specifically, the only difference between these two stories is that in the second one Mary can see patch **A**. Therefore, it has to be the case that seeing patch **A** has a role to play in Mary's coming to believe that "**A** and **B** look the same colour". Consequently, and this derives from the *DPM* account of introspection, seeing **A** should enable Mary to have thoughts about the fact that in seeing both **A** and **B** she has the same type of colour experience.

The upholder of the knowledge argument can explain why seeing **A** enables Mary to judge whether or not she has the same type of colour experience. In fact, by seeing **A** she comes to know the *quale* of that type of colour experience. Therefore, when she sees **B**, she recognises that the current colour experience has the same *quale*. However, it seems that the upholder of the ability reply can respond to this explanation.

¹⁰ Here it is assumed that the colour space describes the colours that Mary can discriminate.

¹¹ See the arguments given in section 3.5, at p. 74.

It can be replied that Mary has in fact the same type of scientific beliefs about objects looking the same colour as she had before her release. The only difference lies in the fact that now she can arrive at these beliefs thanks to the abilities she has acquired by having colour experience. Let us illustrate this reply in more detail.

Let us indicate the light composition reflected by patches **A** and **B** with ψ_1 and ψ_2 respectively. Now, if Mary can study **A** and **B** in the laboratory, she will be able to draw this inference:

- (1) Composition of the light stimulus relative to **A** is ψ_1 .
- (2) Composition of the light stimulus relative to **B** is ψ_2 .
- (3) The compositions ψ_1 and ψ_2 determine a colour match.

Therefore, Mary comes to believe and know that:

- (4) **A** and **B** look the same colour.

We now can describe Mary's epistemic progress after her release. In the following, *C* stands for the colour that the patches look to have. On the account under consideration, this is what happens when she leaves the laboratory:

- (5) Mary sees **A** as looking *C*.
- (6) Mary sees **B** as looking *C*.

Therefore, Mary comes to believe that:

- (7) **A** and **B** look the same colour.

The proponent of the ability reply can account for a difference between the situation in which Mary believes (4) and that in which she believes (7). Mary reaches the belief that (4) is the case inferentially by applying logical rules to what she knows about **A** and **B**.¹² On the other hand, the supporter of the ability reply can

¹² By means of certain mathematical procedures involving models called *wavelength mixture spaces*, it is already possible to predict whether two light stimuli are metamers from their wavelength composition. See Clark 1993, pp. 37-42.

maintain that Mary comes to believe (7) in virtue of a recognitional ability. She acquires such ability in seeing **A** and then she exercises it when seeing **B**. Therefore, seeing coloured objects does not give Mary new beliefs about *qualia* that ground a typology of colour experience. The only difference is that the post-release beliefs she uses to classify colour experiences are based on a new recognitional capacity.

To recapitulate, it can be shown that, upon her release, Mary can come to judge two colour experiences as being of the same type by judging that objects look the same colour to her. However, Mary could possess such beliefs before her release. The only difference now is that she can arrive at these beliefs in virtue of her newly acquired abilities. We have now to consider whether the upholder of the knowledge argument can escape this objection.

6.4 Resisting the Ability Reply

The previous section showed how a version of the ability reply could be used to deny that Mary acquires new beliefs about colour experiences.

This section considers whether the supporter of the knowledge argument can escape the ability reply.¹³ I will argue that this can be done if one advances a principle for the individuation of beliefs. This principle is required to show that, after seeing colours, Mary's beliefs about colour experiences are not the same as the ones she had before her release.

The upholder of the knowledge argument should establish that Mary acquires new beliefs about objects looking the same colour. Determining whether Mary

¹³ Defenders and opponents of the knowledge argument have attacked the ability reply. Some have criticised the distinction between knowing how and knowing that. See Crane 2001, pp. 94-95. Others have focussed on the ability hypothesis. These criticisms show that the abilities suggested by Lewis and Nemirow are neither necessary nor sufficient for knowing what it is like to have an experience. See Lycan 1995, pp. 244-249; Loar 1990 pp. 607-608; Levin 1985, p. 479; Nida-Rümelin 1995, pp. 235-237; Tye 2000, Conee 1994; Raymont 1999, Papineau 1993, Bigelow and Pargetter 1990, p. 146, Seager 1991, pp. 155-56.

acquires these new beliefs requires specifying a principle of individuation for beliefs. This principle will state under which conditions two beliefs are the same. Moreover, if such conditions do not obtain, we can establish that the beliefs differ.

Usually beliefs are taken to be mental states consisting in having an *attitude* towards a certain *content*. If John believes that the snow is white, John has the attitude of believing toward the content expressed by the sentence “the snow is white”. The difference we are investigating cannot depend on Mary's attitudes.¹⁴ Instead, we are interested in whether it can be maintained that she learns something about objects looking the same colour. Thus, the upholder of the knowledge argument needs an account of the individuation of contents.

Two types of conditions individuate the content of beliefs.¹⁵ In one sense, we can maintain that the content of a belief is individuated by the fact that renders it true. Thus, the belief that London is a town is different from the belief that dogs have tails. In fact, these beliefs are true when different facts obtain.

The upholder of the knowledge argument cannot use this account. In fact, showing that Mary acquires new beliefs would be equivalent to showing that there are new facts that these beliefs are about. However, this would mean to assume the conclusion of the knowledge argument.

However, this way of individuating beliefs seems to fail when we want to explain certain differences between beliefs that depend on what subjects *know* or *believe*. For example, someone might not know that Tully is identical to Cicero. Thus, he can believe that Cicero is an orator without believing that Tully is an orator. It can be suggested that the subject fails to have both beliefs because they

¹⁴ Mary's degree of confidence that two things look the same colour might vary according to the method she has used to reach this conclusion. For example, she might think that seeing coloured objects provides less certain results than using scientific instruments.

¹⁵ The debate on the nature of content intersects with that on the nature of meaning of linguistic expression, which constitutes a central issue in contemporary philosophy. This determines an area of research whose extension and depth cannot be addressed here. Useful introductions to the contemporary discussion on content are given in Peacocke 1994 and Papineau 1994.

differ in their content. In the following, I will investigate how we can individuate contents when they are understood in this latter sense. Differences in the informativeness of beliefs depend on the concepts exercised in having them. Believing that the snow is white requires having the concepts [snow] and [being white] (I use square brackets to designate concepts). Just as the words “snow” and “white” are constituents of the sentence “the snow is white”, the concepts [snow] and [white] are the constituents of the content expressed by that sentence. Two contents differ when they have different constituents. Therefore, differences between the content of beliefs are based on differences between the concepts that constitute these contents.

These intuitions concerning how concepts determine the cognitive content of beliefs can be expressed precisely. Christopher Peacocke provides the following principle for the distinctness of concepts:

Concepts *C* and *D* are distinct if and only if there are two complete propositional contents that differ at most in that one contains *C* substituted in one or more places for *D*, and one which is potentially informative while the other it is not. (Peacocke 1992: 2)

Clearly, this principle individuates concepts in terms of what a subject might find informative. If Tully and Cicero are the same person, then someone might not know this fact and thus he will find it informative that “Tully is Cicero”. That “Tully is Tully” is not informative in this way. Let us now consider how this criterion is relevant for our discussion of Mary's case.

The upholder of the knowledge argument has to determine whether the concept of looking the same colour that occurs in Mary's belief before and after release is the same. Let us consider someone who lacks scientific knowledge of colour vision. It seems that this person can determine when two patches look the same colour and believe that “**A** and **B** look the same colour”. Now the concept of having the same

colour which Mary uses before her release is the concept of the relation two stimuli have when their physical features, like reflectance, determine, in certain observers in certain conditions, similar discriminatory responses. Let us name this concept with the expression [looking the same colour]_s where the subscript *s* stands for “scientific”. It seems that the person who has no idea of the physical properties of **A** and **B**, cannot believe that **A** and **B** *look-same-colour*_s. Let us consider the propositional content expressed by:

(1) **A** and **B** look the same colour if and only if **A** and **B** look the same colour.

Substituting in (1) an occurrence of the concept [looking the same colour] with one of the concept [looking the same colour]_s, the resulting content is informative for this subject. Given the criterion of distinctness of concepts introduced above, the two concepts of looking the same colour differ. Consequently, given that concepts determine the cognitive content of beliefs, the beliefs where these concepts occur also differ.

Thus, the upholder of the knowledge argument can argue that the notions of looking the same colour, available to Mary before and after her release differ. These notions play a fundamental role respectively in the scientific and in the introspective categorisation of colour experiences. In scientific terms, types of colour experience are individuated in terms of the notion of looking the same colour based on statistical elaboration of discriminatory responses to physical stimuli. On the other hand, the upholder of the knowledge argument can argue that the ordinary way to type colour experience is based on different beliefs about objects looking the same colour.

Therefore, the supporter of the knowledge argument can argue that the best explanation of why Mary has these new beliefs about looking the same colour, is that when Mary sees a coloured object she acquires a new belief about the colour

she is seeing. Thus, her belief that “the object looks *this* colour” would give her information about the object that she sees. This is information that she could not acquire before her release.

To sum up, it has been argued that the upholder of the knowledge argument has a resource against the ability reply. She can claim that there is a difference between the propositional knowledge that Mary possesses about objects looking the same colour before and after her release. I showed how she can base this difference on a difference in the concepts of looking the same colour. However, if this is the case, it seems plausible that Mary acquires new beliefs about the colours object look to have. Therefore, she acquires new beliefs about the type of colour experiences that she has.

Thus, the upholder of the knowledge argument can rehabilitate a central assumption in his reasoning. When Mary sees coloured objects, she acquires new beliefs and so her knowledge is new propositional knowledge that was not available to her before her release.

6.5 The “Two Ways of Thinking” Reply

I have conceded to the upholder of the knowledge argument that when Mary sees colours she acquires new true beliefs about her colour experiences. Therefore, while Mary is still in the laboratory, she cannot infer these beliefs from her complete scientific knowledge. This section begins to examine whether this assumption implies the ontological conclusion that there are facts that Mary’s scientific knowledge cannot accommodate.

The first aim of this section is to consider one of the implicit premises within the knowledge argument that leads from the assumption that Mary has new knowledge to the conclusion that there are non-physical *qualia*. This is the assumption that necessarily different true beliefs are made true by different facts. The second aim of this section is to endorse the *two ways of thinking reply* to the

knowledge argument that targets this principle. The main idea of this objection is that knowing what it is like to have a colour experience is just a new way to know facts that Mary already knew while she was in the black and white laboratory. By endorsing this reply, I will argue that Jackson's ontological conclusion does not follow from the claim that Mary acquires new true beliefs about colour experiences.¹⁶

The inference from the assumption that Mary acquires new knowledge about colour experience to the conclusion that there are non-physical *qualia* is based on three deductive steps. The first of these steps leads from the premise that Mary has a new true belief about what it is like to have a certain colour experience to the conclusion that *there is* a fact involving the occurrence of a *quale*. The second deductive step shows that this fact is a *new* fact that Mary could not know before her release. The third and final step aims to prove that the *quale* involved in this fact *cannot be a physical property*.

I investigated the premises involved in the first and third inferential step in the previous chapters.¹⁷ Here I recall them briefly. The first step involves some principle that bridges true beliefs and facts. I argued that both a conception of the content of beliefs, or the notion of truth as correspondence, might provide such a principle. The third inferential step is justified by the conception of physical fact involved in the knowledge argument. According to this principle, if a fact about colour experiences is physical then it is known or knowable by Mary. Therefore, if there are facts about colour experiences that she does not know while in the black and white room, these facts cannot be physical. Let us now consider the remaining inferential step.

¹⁶ Some philosophers, instead of considering facts, distinguish between two types of information; see, for instance, Horgan 1984 and Lycan 1995. Although this might reflect their different approaches to the content of beliefs, these authors use a strategy similar to that used here, which is articulated in terms of facts.

¹⁷ See, for the first step, Chapter 4, p. 88. For the third step, see Chapter 2, p. 47.

If it is assumed that different beliefs are made true by different facts, then the ascription to Mary of a new belief about her colour experience implies the existence of a new fact. The reasons for this are as follows. From the premise that Mary acquires new knowledge, and thus a new true belief, we have to conclude that there is a *new* fact she did not know before her release. Given that by seeing a coloured object Mary has a true belief, there should be a fact that makes this belief true. Now, we have to establish grounds for the conclusion that this is fact is a new one that Mary did not know about before her release. The assumption that Mary has a *new* belief appears to provide such grounding. Clearly, this requires the following principle for the individuation of facts:

- (1) Facts F_1 and F_2 are distinct if they make true distinct beliefs B_1 and B_2 respectively.

The discussion of this principle is at the core of two ways of thinking reply. In particular, this objection's starting point consists in revealing that the notion of 'fact' in (1) is ambiguous.

The two ways of thinking reply to the knowledge argument hinges on the distinction between two ways of understanding facts.¹⁸ According to one reading, a necessary requirement for establishing the identity of facts is the identity of concepts used to represent these facts. For example, Peter Carruthers calls this a *thin* notion of fact:

Facts in this sense are just true thoughts, or the mirror-images of true propositions; and they differ whenever the concepts out of which those

¹⁸ In certain formulations of the two ways of thinking reply, this distinction concerns different types of belief contents or kinds of information. These terminological differences might reveal deeper theoretical divergences. However, these issues appear not to be relevant at the level of generality of this presentation.

thoughts or propositions are built are different from one another.
(Carruthers 2000: 33)¹⁹

The individuation conditions of thin facts depend on the ways in which we think about their constituents. For example, the thought that “the glass contains H₂O” and the thought that “the glass contains water” involve different ways of representing a state of affairs in the world. While the former thought involves the concept [H₂O] the other involves the concept [water]. Therefore, the thin facts associated with these thoughts differ.

There is an understanding of facts alternative to the *thin* notion. Following again Carruthers, we can call this a *thick* notion of fact.²⁰ This way of thinking about facts is used:

... to characterise what is there in the world considered as distinct from our modes of representation of it. Here one and the same fact can be represented by many different thoughts. (Carruthers 2000: 33)

This understanding of facts stresses their independence of certain ways of conceptualisation. In order to distinguish thick facts it is not sufficient to determine that we have beliefs about these facts that involve different ways of thinking. In particular, different beliefs can be about the same thick fact. Consider, for instance, the belief that “the glass is full of water” and the belief that “the glass is full of H₂O”. Arguably, these beliefs involve different thoughts that represent the same thick fact.

¹⁹ Michael Tye, in offering his version of the two ways of thinking reply, advances a similar account for what he calls “fine-grained facts”: “Facts are sometime taken to be as fine grained in their individuation conditions as the contents of the propositional attitudes. [...] What distinguishes these facts are the different conceptual modes of representation they incorporate.” (Tye 1995: 172)

²⁰ Similarly, Tye maintains that: “There is another, more coarse grained view, of facts that identifies them outright with states of affairs that obtain in the objective world, regardless of how those states of affairs are conceived.” (Tye 1995: 173)

The distinction between thin and thick facts might be regarded as depending on a distinction between different types of properties. In general, facts can be taken to be structured entities that involve the instantiation of a property in a certain object. Therefore, individuating facts requires as a necessary condition individuating the properties whose instantiations constitute them. Using Peter Carruthers's terminology, we can distinguish between *thick* and *thin* conceptions of properties.²¹ According to the thin conception, properties are individuated as finely as the concepts that refer to them. Thus, conceptual differences determine differences in the relative thin properties. On the other hand, thick properties are not so individuated. For example, the concept [H₂O] and [water] pick out the same thick property – that of being H₂O. However, there is a difference between the thin properties “being water” and “being H₂O”. Let us see how the distinctions between thin and thick facts and properties apply to Mary's case.

Mary comes to know about new thin facts concerning the colours objects look to have. In the previous chapter, I argued that Mary, upon release, acquires new true beliefs about the colours that objects look to have. Therefore, by seeing a red object she might come to acquire the new true belief:

(2) This object looks red.

The concept [red] contained in the thought expressed by (2) differs from any of the scientific concepts that Mary could use in her room. This implies that Mary's new concept individuates a thin property of red, and that there is a new thin fact that Mary comes to know. This has consequences for Mary's knowledge of colour experiences.

Mary comes to know about a new thin fact concerning her experience. The ascription (or self-ascription) of a type of experience is based on the ability to have

²¹ Carruthers 2000, p. 35.

thoughts about the colours objects look to have. In particular, this derives from the account of introspection as a form of displaced perception defended in the previous chapters. Therefore, once Mary believes (2) by seeing a coloured object she can conclude that:

(3) I am having the type of colour experience one has when something looks red to her.

Having such a belief means that she can conceptualise a property of her experience in virtue of a description that involves the concept [red]. Thus, she comes to know a thin fact that was not available to her before her release.

The existence of thin facts about experiences is not problematic for physicalism. There is no reason to assume that Mary should know all the thin facts about colour experience before her release. Thin facts are determined by the ways in which she can conceptualise her colour experiences. However, there is no requirement that a complete scientific knowledge of colour or colour vision should provide knowledge of all the possible ways of conceptualising the phenomena that this knowledge describes and explains. Therefore, proving the existence of thin facts concerning experience is weaker than the conclusion the upholder of the knowledge argument intends to prove. Let us consider how to characterise this stronger conclusion.

The supporter of the knowledge argument must prove that Mary comes to know new thick facts about colour experiences. The following reasons can show this point. Let us assume that by seeing colours Mary acquires a new true belief about a colour experience made true by a new thick fact. Then, she comes to know about a new fact concerning a colour experience that is independent of the way of thinking about it that she might acquire by seeing colours. If a fact concerns colour experiences and is knowable independently of seeing colours, Mary should know it before her release. In fact, in the knowledge argument it is assumed that Mary's

scientific knowledge concerns all the facts concerning colour experiences that can be known independently of having these experiences.²²

The existence of new facts cannot follow logically from our attribution to Mary of new beliefs about her colour experiences. This ascription of new beliefs to Mary is based on the conclusion that she acquires new concepts about her experiences. However, this is consistent with the possibility that these new beliefs concern old thick facts she knew before her release.

To sum up, the *two ways of thinking* reply shows that the knowledge argument is unsound. While it is true that Mary acquires new true beliefs about her experiences, it is possible that it is false that these belief concern non-physical facts. For we cannot exclude that Mary's new beliefs concerns thick facts she knew before her release. In particular, we cannot exclude the possibility that *qualia* are physical properties.

However, the viability of this reply to the knowledge argument requires tackling two issues. First, we have to make clear which are the old thick facts that Mary comes to know in a new way after her release. This issue will be considered in the next chapter in section 7.5.²³ In fact, now we have to deal with a more pressing difficulty. The second problem is providing a detailed account of how Mary's new beliefs about colour experiences constitute a new way of thinking about colour experiences. In particular, this requires an account of the concepts that figure in these new beliefs. The next section will investigate this second requirement.

6.6 Requirements on Phenomenal Concepts

We have seen that our ascription of certain new beliefs to Mary upon her release might be consistent with the falsity of the knowledge argument's ontological

²² In the present context, there is no need to establish whether Mary's scientific knowledge concerns facts that are independent of *any* conceptualisation. Instead, the problem is to establish whether there are facts about colour experiences, that vision science cannot conceptualise.

²³ See at p. 181.

conclusion. This results from adopting the *two ways of thinking* reply. This reply has to account for the truth of the following claims. First, Mary cannot know what it is like to have a colour experience before her release. Second, the existence of this knowledge does not imply the existence of non-physical facts or properties. We have seen that this reply explains both phenomena with the assumption of a conceptual difference that is not mirrored at the ontological level. Thus, a worked out version of this response requires an account of Mary's new concepts about colour experiences.

This section illustrates that the two ways of thinking strategy faces a problem. This difficulty emerges when we consider the requirements that this strategy places on Mary's new concepts about colour experiences. Adopting a now customary expression, I call these *phenomenal concepts*.²⁴ I will show that the two ways of thinking reply is viable only if phenomenal concepts satisfy certain requirements. However, it can be objected that assuming the existence of concepts with these features is an *ad hoc* manoeuvre. What is required is some independent justification for the idea that these concepts satisfy these requirements.

The two ways of thinking strategy places important conditions upon phenomenal concepts. First, possessing phenomenal concepts requires having colour experiences. While Mary is still in the laboratory, she refers to the colour that an object looks to have to a certain individual *S* in condition *C* as a certain position in the colour space. Let us assume that she uses a certain expression *P* as a shortcut for such a description. She can describe the experience that *S* is undergoing as follows:

²⁴ Recent versions of this strategy are pursued by Carruthers 2000, Horgan 1984, Lycan 1995, Rey 1992, Tye 2000, Papineau 2002, Sturgeon 2000, Perry 2001, Loar 1990.

(1) *S* has the type of colour experience that individuals have when things look *P* to them.²⁵

The concept [P]_s, that occurs in the content of belief (1) differs from the concepts [P] she acquires by experiencing coloured objects. Consequently, her concepts about type of colour experiences will differ. As Papineau puts it, the concept of experience she will acquire by seeing a coloured experience will be:

Ways of referring to conscious experiences which are standardly available to human beings only after they have actually undergone the experience in question. Their possession is standardly consequent upon some earlier version of the type of experience they refer to. (Papineau 2002: 96)

Let us consider a second requirement.

Phenomenal concepts are not *a priori* connected to scientific concepts of colour experiences. A concept *A* is *a priori* connected to a concept *B* when, if a subject *S* knows that *A* (*a*), then, without further evidence provided by experience, *S* can come to know that *B* (*a*). For example, if I know that John is the husband of Mary, then I can know without further evidence that John is married. Thus, the concepts of being a husband and being married are *a priori* connected. However, the phenomenal concepts [red] and the scientific concept [red]_s are not *a priori* connected. In fact, we have seen that Mary, from knowing while she is in the laboratory that a certain object looks [red]_s to someone, she cannot infer that the object looks [red] to that person. Similarly, although while she is still in the laboratory she can know that other individuals have the type of colour experience one has when something look [red]_s to him, she cannot know that they have the type of experience in virtue of which things look [red] to them.

²⁵ Of course, she might even know that *S* has a brain state. This is the state that realises the sort of experience that individuals have when something looks *P* to them.

An objection to physicalism, considered by J.C.C Smart, illustrates the importance of another requirement on phenomenal concepts.²⁶ He maintained that experiences are identical to brain states.²⁷ He coupled this metaphysical claim with the assumption that statements reporting experiences, although they involve reference to brain states, have a different meaning from the sentences used in neuroscience. Clearly, this appears to be a central intuition that operates in the new way of thinking reply. Amongst many objections to his type-identity theory, Smart reports one offered by Max Black.²⁸

Black observed that two co-referential expressions provide different ways of thinking about their referent because they characterise it as having different properties. Smart illustrates this point with an example:

For suppose we identify the Morning Star with the Evening Star. Then there must be some property, which logically imply that of being the Morning Star, and quite distinct properties which entail that of being the Evening Star. (Smart 1959: 172)

While the expression “Evening Star” presents its referent as having the property of being visible on the evening, the “Morning Star” involves the property of being visible in the morning. Thus, we can concede that expressions about experiences are co-referential with certain neuroscientific ones. However, according to Black, if these expressions differ in their modes of presentation, then this must mean that there are certain irreducible mental properties. It is in virtue of these properties that

²⁶ Smart 1959.

²⁷ Many of the promoters of this doctrine combined the thesis that type of experiences are identical with type of brain states with the assumption that ordinary talk and knowledge of experience involves a way of thinking about physical states of the brain. See Place 1956, Smart 1959, Feigl 1967. Feigl maintained that this idea should be traced further back to Schlick and Spinoza. He also claimed that type-identity theorists could provide a better formulation by using Frege’s distinction between the *sense* and the *reference* of an expression. See Feigl 1960.

²⁸ Smart 1959.

we can think about brain states as experiences. Nevertheless, this means that we have to accept a form of property dualism.

Smart's reply to this objection cannot help us. He assumed that the descriptions associated with ordinary concepts of experiences involve *topic-neutral descriptions*.²⁹ Broadly, these descriptions refer to experiences by means of causal role specifications. Clearly, Mary can possess these concepts before her release. Thus, if we want to endorse the two ways of thinking reply, we have to provide an account of how phenomenal concepts offer ways of thinking about physical properties without incurring Black's objection.

The plausibility of the two ways of thinking reply requires that phenomenal concepts should satisfy certain conditions. However, do we have any independent and adequate reason for endorsing such a view on phenomenal concepts? So far, I have not provided one. Thus, there is a legitimate challenge to the two ways of thinking reply to the knowledge argument.³⁰

To recapitulate, the viability of the two ways of thinking reply requires more than showing that it blocks the knowledge argument. What is required is some deeper explanation of the special features enjoyed by phenomenal concepts. First, we need to explain how having colour experience provides Mary with concepts of colour experience that are not physical or functional. Second, we have to explain how Mary's new concepts about colours provide new ways of thinking about properties, considered by her scientific knowledge, without referring to any non-physical property.

6.7 Conclusion

In this chapter, I argued that there is a plausible way to resist the ability reply. Thus, upon her release, Mary acquires new true beliefs about the type of colour

²⁹ Smart 1959.

³⁰ Some authors think that this challenge cannot be met; see, for instance, Levine 2001, p. 86.

experiences that she is having. New concepts about colours figure in the content of these new beliefs. However, conceding that, upon her release, Mary forms new beliefs about colour experiences does not imply that there are facts and properties that escape her scientific knowledge.

Endorsing the two ways of thinking strategy counters this implication. However, this strategy requires explaining why phenomenal concepts satisfy certain conditions. The next chapter considers whether this challenge can be met.

7 Different Ways of Thinking About Colour Experiences

7.1 Introduction

The knowledge argument is based on a central inferential step. The premise of this deduction is that Mary acquires new propositional knowledge about what it is like to have a colour experience. The conclusion is that there are facts involving the occurrence of *qualia* that her scientific knowledge cannot accommodate.

This inference can be opposed by denying its premise. We have discussed two ways of formulating this type of criticism. The second chapter examined an argument promoted by Patricia Churchland and Daniel Dennett. According to them, Mary might imagine, or otherwise figure out, what colour experiences are like before her release. The previous chapter presented a second way of *deflating* Mary's putative knowledge of what it is like to have an experience. David Lewis and Laurence Nemirow maintain that, by having colour experiences, Mary acquires a form of knowing how. This is not knowledge of facts. Hence, it cannot be knowledge of *new* facts that Mary did not know before her release.

In the previous Chapter, against these deflationary attempts, I argued that by seeing colours Mary acquires new true beliefs about what it is like to have colour experiences. Therefore, a premise in the main inferential step of the knowledge argument is defensible. However, we should not conclude that there are facts about colour experiences that escape Mary's scientific knowledge. As we saw, the *two ways of thinking reply* blocks this inference. Nevertheless, the plausibility of this reply requires an account of phenomenal concepts. These are the concepts that, supposedly, enable Mary to have new beliefs about colour experiences when she finally sees coloured objects.

The present chapter considers whether there is an account of phenomenal concepts that can support the two ways of thinking reply to the knowledge argument. The main conclusion of this chapter is that the existence of recognitional colour concepts provides this grounding. Once Mary sees coloured objects, she acquires recognitional colour concepts. Having these concepts enables her to think in new ways about the different types of colour experiences. However, these thoughts concern facts she already knew before her release.

Section 7.2 presents a very articulated formulation of the *two ways of thinking* reply offered by John Perry. This account is based on the assumption that Mary's new knowledge is a form of indexical knowledge based on the use of the demonstrative concept [this]. Perry's account appears to match the requirements for a viable version of the *two ways of thinking* reply. However, in section 7.3, I argue that this account does not succeed insofar as it does not cover all what Mary comes to learn by having colour experiences. Section 7.4 presents what seems to be a more plausible version of the two ways of thinking reply. The central assumption in this account is that our colour concepts are recognitional. Finally, section 7.5 motivates the idea that Mary's new ways of thinking about colour experience do not require the existence of properties of colour experiences that escape her scientific knowledge.

7.2 The Indexical Reply

In the previous chapter, we saw that the viability of the two ways of thinking reply requires explaining certain features of phenomenal concepts. These concepts, that refer to colour experiences, should have the following characteristics. First, they cannot be reducible to the scientific concepts (physical or functional) available to Mary while she is in the laboratory. Second, possessing these concepts requires having colour experiences. Third, phenomenal concepts provide new modes of thinking about physical entities without involving non-physical properties. Thus, we

have to investigate whether the upholder of the *two ways of knowing* reply can provide a satisfactory account of phenomenal concepts.

Some philosophers have offered a version of the *two ways of thinking* response that can be called the *indexical reply*. They maintain that Mary's knowledge of what it is like to have a colour experience is best analysed as a form of indexical knowledge.¹ This is knowledge that involves beliefs whose content is expressible by sentences containing *indexical terms* such as "I", "here", "now", "this" and "that". The shared feature of these terms is that their referent changes according to their context of use. For instance, the reference of the term "I" depends on who uses this expression. For example, if John says "I am glad", "I" will refer to him, if Mary says "I am glad", "I" will refer to her. Similarly, the demonstrative "this" will refer to the object attended to by the user.

The aim of this section is to illustrate the indexical reply as recently elaborated by John Perry.² His response to the knowledge argument is based on an articulated theory of the content of belief expressible with sentences involving indexical expressions. I will proceed as follows. First, I will illustrate the intuitive source for the assumption that knowing what it is like to have a colour experience is indexical knowledge. Second, I will consider the details of Perry's indexical reply.

Perry thinks that realising that knowing what it is like to have an experience is indexical knowledge leads to a satisfactory formulation of the *two ways of thinking* reply. The upholder of this reply has to account for two main phenomena. First, she has to explain why, while Mary is still in the laboratory, she cannot deduce, or otherwise figure out in virtue of her scientific knowledge, what it is like to have a colour experience. Second, it has to be shown why this inability does not imply the

¹ An explicit identification of knowing what it is like to have an experience and indexical knowledge is presented in Bigelow and Pargetter 1990, Horgan 1984, Ismael 1999, Perry 2001, Rey 1992. See also McMullen 1985 where an exploration of the analogies between these two forms of knowledge leads to a physicalist response to the knowledge argument.

² Perry 2001.

existence of non-physical facts. Let us consider an example to illustrate how indexical analysis might account for these phenomena.

John is transported, while asleep, to a windowless room in Edinburgh.³ Therefore, he does not know his location. However, he has a map of Scotland that gives him detailed information about Edinburgh and its spatial relations with the rest of the country. What he knows by means of the map can be specified completely in terms of beliefs involving descriptions of distances and the relative positions of different towns, villages and other places in Scotland. For example, he might know that:

(1) Edinburgh is south of Inverness.

Despite the accuracy of his map, it seems that John lacks certain knowledge about the position of Edinburgh before leaving the room.

While John is still in the room, there is some indexical knowledge he cannot deduce from the knowledge provided by the map. For example, he cannot come to know:

(2) This place is Edinburgh.

Similarly, he cannot know:

(3) This place is south of Inverness.

No matter how detailed the map; if it contains only information expressible by means of non-indexical expressions it cannot inform John about his position.⁴ Only by escaping and recognising the town where he is, can he know (2) and (3).

Although John cannot come to believe (3) from the beliefs he can acquire by consulting the map while he is still in the room, it seems that the map does not leave

³ See other examples in Perry 2001, 101-113.

⁴ Assuming that the map represents John's position, by means of the correct "you are here dot", would beg the question.

out any fact about the position of Edinburgh in Scotland. John's knowledge of (3) does not appear to involve a new fact about Edinburgh's position he could not know before his release. In fact, when he was still in the room he knew (1) because the map contains the information that Edinburgh is south of Inverness. Intuitively, it appears that by leaving the room and recognising the town where he is, John can come to think about this fact in a new way. John's example appears to have some similarity with Mary's case.

Mary's scientific knowledge comprises beliefs expressible by means of sentences containing scientific descriptions of colour experiences that she can read about in her books or screens. Now, as in John's case, if knowing what it is like to have an experience is indexical knowledge, Mary cannot possess it before her release. In particular, it might be plausible to assume that her new knowledge does not concern non-physical facts. Let us now consider in detail how Perry's formulates his indexical reply.

According to Perry, before her release Mary can know to be true statements of the type:

(4) Q_r is what it is like to see red.

He introduces Q_r as an expression that Mary uses to refer to what it is like to see a certain colour while she is still in the laboratory. Given that he endorses a version of the two ways of thinking reply, this is a perfectly legitimate assumption. Q_r refers to a physical state or a physical aspect of the normal experience of seeing red. On the other hand, when Mary comes out of her laboratory and sees a red object of which she knows the colour, she acquires a belief whose content is expressed by:⁵

⁵ Perry claims that Mary can recognise the object by the shape and know that type of object has a certain colour. For example there is no reason to deny that before her release she might know that tomatoes are red and that she knows how to recognise tomatoes from their shape. Perry 2001, p. 99-100.

(5) *This_i* is what it is like to see red.

The demonstrative *this_i* is related to Mary's ability to make epistemic contact with a feature of her colour experience, that he calls subjective character, in a subjective way. Perry illustrates this point as follows:

When we are attending to a subjective character in the subjective way and wish to communicate what we are feeling or noticing, we use our flexible demonstrative, "this", as in "this feeling is the one I've been having. Let's label this use of "this" as an inner demonstrative: "*this_i*".
(Perry 2001: 146)

Once Mary knows that (5), she can then discover that:

(6) *Q_r* is *this_i* subjective character.

Therefore, Perry's main task is to show that although (6) is informative for Mary, this does not imply that she comes to know about a non-physical property.

Perry's response to the knowledge argument follows the general strategy of the *two ways of thinking* reply. First, he draws a distinction between different types of facts or contents related to beliefs.⁶ Second, he uses this distinction to explain why Mary cannot know what it is like to have a colour experience without implying the existence of non-physical facts. With respect of the first stage of Perry's reply, the crucial distinction is between the *subject matter* and *reflexive content of a belief*. Let us clarify these two notions of content.

Perry assumes that beliefs are structured representations obtained by the combination of more basic representations he calls *ideas*.⁷ Amongst these latter representations, *notions* stand for individuals such as particular persons or towns. On the other hand, *concepts* refer to universals such as properties and relations. This

⁶ Perry 2001, p. 113.

⁷ Perry 2001, p. 50-51.

representational account of belief is coupled with a causal account of the reference of the constituents of beliefs. An idea stands for the entity that is causally its *origin*.⁸

According to Perry, the content of a belief is assigned under the assumption that certain conditions about the belief are satisfied. In fact, the content of a belief is given by what renders the belief true. Moreover, the assignment of these truth conditions to a belief is relative to those facts about the belief we take as given. In particular, these are facts about the ideas and notions that figure in these beliefs. The dependence of the content of a belief on certain conditions concerning that belief can be illustrated with an example. Let us consider the belief b_1 that:

(1) Edinburgh is south of Inverness

Now, b_1 is true when Edinburgh is south of Inverness and this latter fact is the content of John's belief. The assignment of content to belief b_1 depends in part on facts about this belief that we take as given. For example, it is assumed that the notion "Edinburgh" and the concept "being south of Inverness" refer respectively to Edinburgh and to the relation of being south of Inverness.

Perry illustrates the dependence of a certain belief's content on certain contextual facts about the same belief by means of a formula he calls the "content analyzer".⁹ The *content analyser* characterises the content of a belief by specifying the truth conditions of that belief, given certain facts about the belief. The general form of this formula is:

CA: Given *such and such*, ϕ is true iff *so and so*.

⁸ Perry 2001, pp. 51-53.

⁹ Perry 2001, p. 125.

The letter ϕ stands for truth-valuable representations such as cognitive states like beliefs or, more generally, thoughts.¹⁰ While the expression “*such and such*” stands for contextual facts about the representation ϕ , “*so and so*” stands for the content assigned to ϕ once these facts are given. Having put into place the details of Perry's proposal, let us now consider the distinction he draws between subject matter and reflexive content.

The subject matter is the content of a belief that is assigned when all the facts about the representations involved in the belief are given. We can characterise this content by means of the content analyser. Let us consider the belief b_1 . Once all the facts about the ideas involved in b_1 are specified, we obtain this instance of the content analyser:

Given that the idea [Edinburgh] is of Edinburgh and the concept [being south of Inverness] is of the relation being south of Inverness, b_1 is true iff *Edinburgh is south of Inverness*.

The subject matter of b_1 is given in Italics. In particular, we can think about the subject matter of a belief as the thick fact that renders the belief true. The characterisation of this fact does not involve the way in which the subject might come to think about it.

The reflexive content of a belief is individuated when how the ideas that figure in the belief pick out their referents is not assumed as given. Let us use the content analyser to reveal the reflexive content of b_1 . When facts concerning how b_1 is related to its subject matter are not given, the condition “*such and such*” in the content analyser is empty. Therefore, we obtain that:

¹⁰ Perry applies his analysis also to linguistic expressions, however given that his discussion of Mary's cases is entirely focussed on beliefs, I will consider only beliefs.

b_1 is true iff *the notion [Edinburgh] in b_1 is of Edinburgh and the concept [being south of Inverness] in b_1 is of the relation being south of Inverness and Edinburgh is south of Inverness.*

Now the content of b_1 , given in Italics, involves conditions concerning the belief itself, i.e. about the ideas that constitute it. This explicit reference to the belief explains why this content is named reflexive. Clearly, the reflexive content involves not just the fact that Edinburgh is south of Inverness but equally mentions the ideas that constitute the belief. In particular, the requirement that the conditions that make an idea stand for a certain entity are satisfied is itself made explicit.

Perry maintains that two beliefs can have the same subject matter and different reflexive contents.¹¹ In particular, this might happen when one belief involves indexical representation and the other does not. This case can be illustrated by considering John's beliefs. We have already seen the subject matter of belief b_1 . Let us now consider the subject matter of John's belief b_3 whose content is expressible by the sentence:

(3) This place is south of Inverness.

According to Perry, determining the subject matter of a belief such as b_3 requires that we take as given a relation of *attachment* between this belief and John's visual perception.¹² In fact, b_3 is related to John's perceiving a certain place. The subject matter of this belief depends on which place he is perceptually attending to. Thus, the content analyser has the following form:

Given that the perception attached to b_3 is of Edinburgh and the concept [being south of Inverness] in b_3 is of the relation being south of Inverness, b_3 is true iff *Edinburgh is south of Inverness.*

¹¹ Perry 2001, p. 113.

¹² Perry 2001, pp. 120-121.

The subject matter of belief b_1 and b_3 is that Edinburgh is south of Inverness. However, their reflexive contents differ. The reflexive content of b_3 is given in Italics in the following instance of the content analyser:

b_3 is true iff *the perception attached to b_3 is of Edinburgh and the concept [being south of Inverness] in b_3 is of the relation being south of Inverness and Edinburgh is south of Inverness.*

The reflexive content of b_1 and b_3 involves references to different entities. In particular, the reflexive conditions of b_1 do not mention the perception involved in John's attending to Edinburgh that is involved in the reflexive content of b_3 .

According to Perry, the *reflexive content* of beliefs is central to an understanding of the roles that beliefs play in motivating subjects' inferences (or actions). Thus, reflexive content is relevant for the cognitive roles of beliefs. It is important to notice that Perry recognises that reflexive content does not explain directly why individuals can have different attitudes towards the same subject matter. Subjects do not usually have explicit beliefs about the reflexive truth conditions of their beliefs.¹³ In particular, we can exclude the suggestion that explicit knowledge of reflexive contents explains why someone might have different attitudes to thoughts about the same subject matter.

To illustrate this point, let us consider again John's case. We can assume that he has never heard about states of perceptual attention, or about linguistic notions such as "term" or semantic notions as "reference". Still, it makes sense to say that before his release he believes (1) without believing (3). But, as Perry points out:

Nevertheless, the reflexive content, the truth-conditions our content analyser give us when we do not take the contextual facts as given, is

¹³ Perry 2001, p. 132, p. 135.

essential in understanding the motivation for making the statements and what is involved in understanding them. (Perry 2001: 129)

Thus, we have to consider what kind of account of the cognitive role of beliefs this content provides.

According to Perry, explicating the reflexive content reveals the different ways in which certain expressions or certain mental representations refer to certain objects. In particular, the difference between notions that are “*detached*” and “*attached*” to perception is revealed.¹⁴ Perry would maintain that the notion [Edinburgh] that John has in the room is about Edinburgh because there is a certain causal relation that renders Edinburgh the origin of that notion.¹⁵ For instance, we can imagine that John’s notion [Edinburgh] is linked to Edinburgh by a series of causal connections. This causal chain might involve the effects on the cartographer that determined the original drawing of the map, and John’s use of the map. Thus, John is able to think about Edinburgh in virtue to this causal connection. However, his notion is not attached to his current perception of Edinburgh. On the other hand, certain notions are attached to this perception. When John leaves the room and comes to believe (2), he attaches his notion [Edinburgh] to the perception that supports his ability to attend to Edinburgh. Now, according to Perry, attaching a certain notion to a perception determines a “flow of information” from the perception to the ideas associated with the notion. It is in virtue of this flow of information that John can come to believe (3). Thus, it seems that Perry’s account has all the resources to explain why a difference between indexical and non-indexical knowledge does not imply an ontological difference at the level of subject matter. We can now consider Mary’s case.

¹⁴ Perry 2001, p. 124.

¹⁵ Perry 2001, p. 51.

Perry argues that the knowledge argument is based on the *subject matter fallacy*. One commits this fallacy when one believes that:

... the content of a statement or a belief consists in the conditions that the truth of the statement or belief puts on the objects and properties the statement or belief is about. (Perry 2001: 20)

However, we have seen that besides the *subject matter* of a belief there is its *reflexive content*. Now, Mary's discovery that:

(6) Q_r is *this*_i subjective character.

is made possible by "a new fact at the level of reflexive content"¹⁶ and not by a fact at the level of subject matter. The belief that (6) is true:

iff the *act of inner attention to which it is attached is of the subjective character of the experience of seeing red*. (Perry 2001: 148)

Before her release, Mary lacked an "informational link" between her "detached" notion of a physical property of the brain and her "attached" notion of the qualitative feature of seeing-red sensations.

Perry's account of Mary's case satisfies the two fundamental requirements of the *two ways of thinking* reply. First, he can explain why she cannot know what it is like to have an experience. This is because her scientific beliefs, and those involved in this knowledge, have different reflexive content. Second, he can argue that such a difference does not logically imply that these beliefs concern different facts as assumed in the knowledge argument. In fact, an indexical belief and a non-indexical one can have same subject matter.

Moreover, Perry's account appears to explain the different requirements that the two ways of thinking strategy places on phenomenal concepts. First, phenomenal concepts are not reducible to Mary's scientific concepts because they are

¹⁶ Perry 2001, p. 159.

demonstratives. According to Perry, Mary's scientific knowledge does not involve such demonstrative concepts. Moreover, indexical concepts cannot be reduced to non-indexical ones. Second, phenomenal concepts require having experiences because it is only by having these mental states that a subject is put in a position to attend to their subjective character and thus form demonstrative concepts and thoughts about it. Finally, it seems that Perry can evade Max Black's objection.

The main upshot of our discussion of this objection was that phenomenal concepts cannot be descriptive.¹⁷ In fact, this would result in a dilemma for the upholder of the *two ways of thinking* reply. Either these phenomenal concepts would be available to Mary before her release, or they would refer to non-physical properties. Perry thinks that his proposal fares very well in this respect, given that he maintains that indexical concepts are not descriptive.¹⁸ Demonstrative expressions refer directly to their referents. However, demonstrative concepts contribute to the cognitive content of the thoughts, where they occur, without involving way of thinking specified by a description.

To recapitulate, it appears that Perry offers a promising and articulated justification of the two ways of thinking reply to the knowledge argument. Specifically, he explains the requirements that this response places on phenomenal concepts by means of an independently motivated account of the content of demonstrative thoughts. The next section investigates whether we should accept Perry's proposal.

7.3 Against the Indexical Reply

In this section, I argue that Perry's indexical account of Mary's knowledge of what it is like to have a colour experience is unsatisfactory. Firstly, Perry bases his

¹⁷ See Section 6.6, at p. 151.

¹⁸ The non-descriptive nature of indexical ways of thinking has a central role in his sustained criticism of Fregean semantics, for he takes it that Fregean modes of presentation involve descriptions. See Perry 1977. This latter statement has been forcefully resisted in Evans 1982.

account on a model of introspection that was rejected in Chapter 4.¹⁹ Secondly, even a reformulation of Perry's indexical reply that avoids this problem has to face a general difficulty. Namely, Mary's new knowledge about colour experiences involves more than indexical knowledge.

The notion of an act of inner attending to a feature of experience is central in Perry's account of Mary's case. According to him, having colour experiences enables her to have beliefs she could not have before her release. These beliefs involve the demonstrative concept *this_I* that refer to the subjective character of experience. This demonstrative identification is based on an act of *inner attending* to the subjective character of the experience. In particular, Mary must be related by this act of inner attending to the subjective character of her experience in order to be able to think about this property in a new way that is not available to her before her release. Therefore, determining the nature of this inner act is of central importance in the evaluation of Perry's proposal.

According to Perry, the act of attending to the subjective character of an experience requires having that experience. Clearly, this explains why Mary lacks demonstrative knowledge about this feature of experience. However, having the experience is not sufficient to ground the demonstrative identification.

This act of attending to the experience is quite different than the experience itself. (Perry 2001: 49)

In fact, experiences have their qualitative character, or what is like to have them, independently of our attending to them. Thus, we can have an experience with a distinctive character even if we do not notice it.²⁰ However, by having colour experiences we are also in a position to enter into a certain epistemic relation with these mental states. In particular: "we can attend to their subjective characters".

¹⁹ See sections 4.4, p. 96, and 4.5, p. 100.

²⁰ Perry 2001, p. 49.

Another psychological state should detect the subjective character of the experience. This state should enable Mary to have certain demonstrative thoughts not available to her in the black-and-white room. Let us consider what account can be provided of this act of inner demonstration.

Perry seems to assume that we have direct access to our experiences and their features. He states that the act of inner demonstration, which is required for the knowledge of what it is like to have a colour experience, is not a perceptual process. Nevertheless, he assumes that, as it happens with the things we perceive:

We can be aware of our experiences. We can attend to their subjective characters. We can pay more or less attention to them. We can mentally demonstrate them (“This sensation...”), and communicate facts about them to others. We can notice what there are like, and anticipate what they will be like. (Perry 2001: 47)

Thus, it seems that Perry assumes that, in particular, the ability to think demonstrative thoughts about the subjective quality of a colour experience depends on attending to the colour experience. However, this appears to create a problem for his account.

There is no such direct access to our experience. In fact, this access would require the capacity to have true demonstrative thoughts about these mental states. However, I argued that neither perception nor introspection offer such a capacity.²¹ Moreover, I have suggested an alternative model of the epistemic access to colour experiences and their features. In accordance with this account, we determine the type of colour experiences we are undergoing by paying attention to the way in which coloured objects look to us. An attempt at avoiding this objection might lead to a reformulation of Perry's proposal.

²¹ See, respectively, See sections 4.4, p. 96, and 4.5, p. 100.

Perry's account can be adapted to Mary's new beliefs about the ways in which coloured objects look to her. Thus, we might think that upon her release Mary discovers that:

(1) Red is this

where "this" picks out the colour that the object she is attending looks to have to her. Moreover, the term "red" expresses the scientific concept [red]_s that Mary uses to refer to the colour red. In this case, Mary's new thoughts about the red-type colour experience derive from what she learn when she comes to believe (1). Specifically, the *Displaced Perception Model (DPM)* of introspection can explain how her new beliefs about the red-type experience can be "parasitic" on the discovery expressed by (1).²² Thus, we have a version of the indexical reply that does not require the capacity to attend introspectively to the properties of colour experiences. However, it seems that even such an account is not satisfactory.

According to our version of Perry's account, on her release Mary acquires a new way of thinking about colours when she can demonstrate them. Specifically, she can attach, by means of demonstrative beliefs, her scientific concepts of colours to her perceptions of these properties. Thus, Mary's having new beliefs about colours derives from her capacity to point with the demonstrative "this" to the colours objects look to have to her. In fact, if the two ways of thinking strategy is correct, then Mary can have demonstrative thoughts about the colours objects look to have before her release. In fact, colours are properties that her scientific knowledge can characterise completely. Thus, for instance, by using a certain instrument, she can form the belief:

(1) Red is this colour.

However, this demonstrative thought is based on her perception of a certain instrument and not on directly seeing a red object. If the indexical reply to the

²² See, section 5.5 at p. 125.

knowledge argument is correct, there should be some difference between this demonstrative way of thinking about red and the one available to her when she sees coloured objects.²³ But how can we capture this difference just by considering Mary's demonstrative thoughts?

As some authors have pointed out, it seems that the indexical reply provides a too "thin" characterisation of the content of Mary's new beliefs.²⁴ The difficulty of the indexical account is that it seems that it requires that when Mary has a certain colour experience she is able to refer to the colour she is seeing as "this quality" without really knowing what she is pointing at. Joseph Levine states this problem clearly; Mary might be using "demonstrative pointers":

...with little substantive conception of what sort of thing we are pointing at – demonstrative arrows shot blindly that refer to whatever they hit. (Levine 2001: 84)

This accusation is clear once we remember a central requirement in the two ways of thinking places on Mary's new concepts. Namely, they cannot be descriptive concepts. Now, interpreting her new demonstrative thought as mere pointing can satisfy this requirement. But it leaves unspecified what Mary comes to know about colours.

Some might reply that we should not consider what is at the "demonstrative end" of Mary's beliefs about colours. Instead, these beliefs are new in virtue of their relation to her *egocentric* perspective.²⁵ Some philosophers have found attractive the idea that indexical thoughts can be framed as subject-centred.²⁶ The central intuition of this idea is that:

²³ See for a similar line of criticism of the idea that phenomenal concepts are demonstratives, Tye 2000, p. 25-26.

²⁴ See Tye 2000, p. 26, Chalmers 2003, Levine 2001.

²⁵ This line of reasoning is put forward by Michael Tye, see Tye 2000, p. 25.

²⁶ See McGinn 1983, Lewis 1979.

Very roughly, we can say that to think of something indexically is to think it in relation to *me*. (McGinn 1983: 17)

In particular, a demonstrative concept such as [this] has a special relation with the person who uses it. If someone can use *this*, then he knows *a priori* that the object he demonstrates is the thing that he is attending to. Thus, it might be assumed that this relation with the indexical concept [I] is central in contributing to the cognitive content of the belief where the demonstrative occurs. Let us illustrate how this might work in the case of Mary.

It might be suggested that Mary's new demonstrative beliefs about colour experience differs from those she had before her release because these latter were detached from her egocentric perspective. When she was in the laboratory, she could not formulate a belief whose cognitive content involves a certain direct relation she has with the colour the object looks to have. In fact, she could not have a thought expressible by:

(2) Red is the property to which I am now directly attending.

Nevertheless, it can be maintained that she can acquire such a belief by seeing a red object. Thus, it can be argued that the only difference is that, now, she can relate herself to the colour of the object. In other words, now, she can think about this property as the one to which she is directly attending. Given this new thought, she can then acquire a thought about the type of experience that she is having. She will come to believe that she has the type of colour experience people have when they attend to the property to which she is attending now. Let us see why this proposal is not satisfactory.

The novelty of Mary's demonstrative beliefs does not depend on their relation with her egocentric perspective. When Mary sees a coloured object, she appears to learn more than the fact that she is attending to some property. One reason for this claim is as follows. When Mary is still in the laboratory, she can formulate

demonstrative thoughts about the way in which colours appear to her. In fact, Jackson concedes that she can see black, white and the shades of grey. Thus, she would be already able to have thoughts relative to her egocentric perspective on certain colours.

There is another reason that shows that Mary's egocentric perspective is not the main ingredient in her new thoughts. When she sees different colours she acquires different new thoughts, but her egocentric perspective is the same. Let us assume, for instance, that, upon her release, Mary sees a blue and a red object. Given that she is informed about the colours of these objects, she comes to believe:

(1) Blue is this colour.

(2) Red is this colour.

Now, according to the two ways of thinking reply, (1) and (2) express discoveries made by Mary. Moreover, it seems intuitively plausible that these are different discoveries. However, in both cases Mary's two thoughts are based on what *she* is attending to. Michael Tye has expressed this clearly with regard to the concepts that Mary acquires:

Each phenomenal concept is thus tied to a particular-experience specific perspective occupied by the possessor of the concept. As the experiences vary, so too do the phenomenal concepts. (Tye 2000: 26)

Thus, it seems that we cannot account for Mary's new knowledge in indexical terms.

To recapitulate, we have seen that Perry's account of Mary's new knowledge is not satisfactory, as it depends on his endorsement of a wrong account of introspection. However, even if we formulate his suggestion without assuming such an account, certain difficulties remain. In particular, it seems that Mary learns something more substantial than just the mere ability to point to a certain property by means of a demonstrative concept.

7.4 Recognitional Concepts

The *two ways of thinking* reply appears to provide a plausible answer to our version of the knowledge argument. However, the tenability of this response depends on an account of phenomenal concepts. These are concepts concerning colour experiences that Mary supposedly acquires by seeing coloured objects. The indexical analysis of the knowledge of what it is like to have a colour experience fails to provide a satisfactory account of Mary's phenomenal concepts. Thus, we have to account for these concepts.

The present section's main thesis is that we can formulate a plausible version of the *two ways of thinking* reply. This account is based on the assumption that ordinary colour concepts are *purely recognitional*. First, I will illustrate the main features of this class of concepts. Second, I will show how the hypothesis that ordinary colour concepts are purely visual recognitional offers a plausible account of Mary's case that is consistent with the requirements of the *two ways of thinking* reply. Finally, I will illustrate how the account of phenomenal concepts here offered differs by other accounts that articulate the two ways of thinking reply by invoking purely recognitional concepts.

A plausible account of phenomenal concepts to support the two ways of thinking reply should satisfy the following requirements. First, phenomenal concepts should not be *a priori* connected to any of the scientific descriptions available to Mary before her release. Only if this is the case will the new introspective beliefs, that Mary forms once she sees coloured objects, contain thoughts she could not have or express when she is in the laboratory.

Second, phenomenal concepts do not constitute *two ways of thinking* in virtue of involving descriptions not available to Mary before her release. I illustrated the reason for this requirement in discussing an objection raised by Max Black to the

type identity theory.²⁷ The upshot of that discussion was a dilemma for the supporter of the two ways of thinking reply if he assumes that those phenomenal concepts are descriptive. Either these descriptions are available to Mary before her release, or they require the existence of properties her scientific knowledge cannot account for.

The third requirement is that Mary can acquire these concepts concerning colour experiences by seeing coloured objects. Finally, these new concepts refer to properties that figure in Mary's scientific account of colour experiences. It seems that assuming that Mary acquires purely recognitional colour concepts can satisfy all these requirements on phenomenal concepts.

It has been observed that certain concepts are recognitional. These concepts can be applied on the basis of perceptual or quasi-perceptual acquaintance with their instances. Thus, someone has a recognitional concept [book], when she is able to judge, on the basis of what she sees, whether what she sees is a book. Moreover, it has been argued that there are concepts that are purely recognitional. These concepts can be characterised as follows.

A concept is purely recognitional when its possession-conditions (in the sense of Peacocke 1992) make no appeal to anything other than such acquaintance. A concept is purely recognitional when nothing in the grasp of that concept, as such, requires its user to apply or to appeal to any other concept or belief. (Carruthers 2002: 4)

Let us now consider how to use this notion in the discussion of Mary's case.

The central thesis defended here is that Mary acquires new recognitional concepts concerning colours. We have seen that Mary might acquire new beliefs about her colour experiences only if she acquires new beliefs about the colours that

²⁷ See Section 6.6, p. 154.

objects look to have. Thus, we might consider whether Mary acquires purely recognitional concepts about the colours that objects look to have. If this is the case, she can also acquire new ways of thinking about the types of colour experiences she is having.

Colour concepts appear to be purely recognitional. Let us consider the description “looking the colour of a rose”. We saw that Mary may know what “looking the colour of ___ ” means and know what “a rose” means. However, she does not know *which* colour is referred to by the phrase “colour of a rose”. Now Mary knows that roses have physical properties such that in normal lighting conditions they elicit from normal observers, who have a mastery of the meaning of “red”, and who are asked to name the colour of the rose, the response, “it is red”. Moreover, she can think about red as the colour that has a certain position in a system of relations of similarity described by the complete colour space.

We have already seen that Mary would not be able to recognise which colour is red upon seeing coloured objects.²⁸ The scientific information she possesses about “being red” does not enable her to pick out which is the colour of the rose. This suggests that the concept of being red is essentially visual-recognitional in that only those who can pick out by sight the colour possess full mastery of the concept.

Now, if the concept of red is purely recognitional, then, before her release, Mary cannot have certain beliefs. These are the beliefs where the recognitional colour concept of red figures. In fact, her scientific knowledge provides her with descriptions of red. In particular, if the concept “red” is essentially visual-recognitional, then Mary acquires a concept of “looking red” that she cannot possess before her release. In fact, it seems that someone cannot possess the concept

²⁸ See Section 3.5, at p. 74.

[looking red] without having the concept [red].²⁹ We saw that, in order to possess the notion of looking red, someone has to be capable of judging that something looks, to him, as a red object would look in certain circumstances.³⁰

If [red] is a purely recognitional concept, then Mary can acquire a new belief about the type of colour experience that she is having. She can discover that she has the type of colour experience people have when something looks red to them. In this case, red is the new purely recognitional concept that she acquires by seeing coloured objects.

It seems that assuming that, by seeing red objects, Mary acquires the purely recognitional concept [red], satisfies the requirements of the two ways of thinking reply. First, it is clear why Mary needs to have a certain type of colour experience in order to acquire a new concept about it. By seeing a coloured object, she acquires a new description about the type of colour experience that she is having. However, this description is not available to her before her release because the purely recognitional concept “red” figures in it.

This point can be illustrated if we consider the recognitional concept of a certain person. Someone can know that John is the father of Henry. Moreover, without having ever met Henry, this person can know Henry by means of a certain description, say as “the owner of a certain car”. However, when the person acquires a recognitional concept of Henry, then he would be able to use a new way to describe the father of Henry. He might use the expression the “father of Henry”, as the father of a person he is able to pick out thanks to his perceptual acquaintance with him.

²⁹ Here I am assuming that the concept [red] is *cognitively prior* to [looking red]. This means that no one could possess the concept [looking red] without possessing the concept [red]. However, this is independent of the question of how understanding the relation between the property red and having the experience that something looks red. See Peacocke 1984, pp. 61-63.

³⁰ See Chapter 5, p. 118.

This version of the two ways of thinking reply differs from other accounts that refer to recognitional concepts. In fact, many authors have assumed Mary's new purely recognitional concepts refer to colour experiences.³¹ On this view, Mary acquires new ways of thinking about colour experiences because she can now recognise, without any mastery of other concepts or having other beliefs, that she is having a certain type of colour experience. Instead, I maintain that the novelty of Mary's thoughts about her colour experiences derives from acquiring new purely recognitional concepts about the colour of objects. According to this doctrine, the ordinary introspective capacity to self ascribe colour experiences requires the capacity to judge the colour objects look to have. Moreover, this introspective capacity requires the mastery of some principle that connects having colour experiences of a certain type to the fact that something is looking to us to have a certain colour.

Thus, the concepts we ordinarily use to refer to types of colour experiences cannot be purely recognitional.³² However, the descriptions that we ordinarily use to self-ascribe types of colour experience involve as one of their constituents purely recognitional colour concepts. It is in virtue of the dependence of the ordinary typology of colour experience on purely recognitional colour concepts, that Mary acquires new ways of thinking about colour experiences.

To sum up, once we assume that Mary acquires recognitional concepts of colours, she then acquires new ways of thinking about her colour experiences. Within these new thoughts figure description of the types of colour experiences. These descriptions involve the occurrence of a visual recognitional concept of colour. However, the existence of these thoughts, that Mary could not have before

³¹ See Tye 2003 and Loar 1990.

³² Despite this, there might be ways in which we learn to use concepts of colour experiences in a recognitional way. For example, a doctor can learn to recognise visually a tumour from an x-ray photograph. However, possessing the notion of tumour requires mastering certain theoretical knowledge.

her release, does not imply that Mary comes to know new facts. She can now think about her colour experiences in a different way. Such a way of thinking is grounded on her possession of purely recognitional colour concepts.

7.5 Old Thick Properties

We have seen that we cannot exclude the possibility that when Mary sees colours she comes to know about facts she already knew before her release. This is enough to reject the knowledge argument. However, it can be asked legitimately which facts Mary comes to know in a new way.

This section aims to answer this question. As a preliminary clarification, I will introduce the notion of thick property. Properties of this type are not individuated in terms of the concepts used to refer to them. I will argue that determining which thick facts Mary comes to know requires establishing which *thick* properties her new concepts of *qualia* refer to. Then, I shall argue that these latter concepts refer to thick properties referred by Mary's scientific concepts.

In order to argue for this conclusion, I will proceed as follows. First, I will argue that the identity of thick properties is a function of the causal powers these properties endow objects with. Second, I will argue that the concepts of *qualia* refer to properties that determine certain causal powers. Finally, I will maintain that there are properties referred to by Mary's scientific concepts whose instantiations determine the same causal powers of *qualia*. Therefore, I conclude that *qualia* are identical to these natural properties.

Establishing which old thick facts Mary comes to know requires determining the properties picked out by the concepts involved in her new true beliefs. This is the case for the following reasons. Thick facts are structured entities that involve the instantiation of a property in a certain object. Therefore, individuating facts requires, as a necessary condition, individuating the properties whose instantiations constitute them.

Endorsing the two ways of thinking reply implies that the properties referred to by Mary's new concepts are *thick*. Using Peter Carruthers's terminology, we can distinguish between *thick* and *thin* conceptions of properties.³³ According to the thin conception, properties are individuated as finely as the concepts that refer to them. Thus, conceptual differences determine differences in the relative thin properties. On the other hand, thick properties are not so individuated. The two ways of thinking reply requires that Mary's new introspective beliefs about colour experiences concern the instantiation of thick properties. For we have determined that Mary acquires new concepts she did not possess before her release. Thus, this difference would determine a difference between the properties to which these concepts refer.

Some philosophers have suggested that thick properties are responsible for the causal powers of objects. Objects have causal powers specified by means of conditional statements connecting circumstances in which the object might be involved and certain effects. For example, stating the following conditional can specify the causal powers of a knife: if its blade is applied to a piece of butter in normal conditions, then the butter will be cut. Many philosophers think that the manifestation of objects' causal powers depends on their possessing properties that ground these powers.³⁴ In particular, it can be maintained that objects have certain causal powers because they have properties that will cause, in the appropriate circumstances, the manifestations of such powers. In this sense, it can be maintained that these properties play a certain causal role.³⁵ For example, we can think that in certain conditions the knife cuts the butter given the fact that it has a certain shape, size, and it is made of steel. In particular, we can say that the

³³ Carruthers 2000, p. 35.

³⁴ See Shoemaker 1980.

³⁵ Lewis 1986, Prior 1985, Prior, Pargetter, and Jackson 1982.

occurrence of these properties is causally responsible for the manifestation of the causal powers of the knife in the different circumstances.

The previous considerations suggest that thick properties are individuated in terms of the causal powers they determine. The general form of this individuation principle is well stated by Sidney Shoemaker:

If under all possible circumstances properties X and Y make the same contributions to the causal powers of the things that have them, X and Y are the same property. (Shoemaker 1980: 234)

This criterion is incompatible with the claim that necessarily properties differ when the concepts used to refer to them differ. Although the concept of being [H₂O] and that of being [water] differ, what causally derives from being [H₂O] derives from being [water]. This way of individuating thick properties gives us a way of establishing when two concepts refer to the same thick property. Two concepts refer to the same thick property when we can judge that the properties to which they refer play the same causal role. Let us turn back to the discussion of Mary's case.

The notion of *quale* that is involved in Mary's introspective beliefs has two main features. The first feature is that a *quale* is a property that determines the type of colour experience she is having. The second is that Mary is not directly aware of the *quale*. In fact, we have seen that in seeing coloured objects Mary can formulate the following belief about her colour experience:

I am having the type of colour experience that one typically has when things look this colour.

While Mary is aware of the ways in which coloured objects look, she is not directly aware of the properties of her experiences. She determines the type of experience she is having in virtue of judgements about the colours that objects look to have. However, it seems that we can provide a more substantial account of the property

enjoyed by the colour experience a subject has when an object looks a certain colour to her.

Qualia are properties of experiences whose occurrence is responsible for subjects' visual discrimination, when stimulated under certain conditions, of the colours that objects look to have. It is in virtue of having a colour experience with a certain *quale* that a certain object looks red to Mary when she is affected by certain stimuli. This means that when Mary thinks introspectively about the property that determines the type of colour experience she has, she refers to a property responsible for a certain causal power.³⁶ This gives a clear indication of what thick properties are referred to by Mary's concepts of *qualia*.

There are scientific concepts that refer to properties that are responsible for the discrimination of colour stimuli. As we saw, psychophysics aims to describe the discriminatory capacities of subjects.³⁷ This discipline determines the dimensions along which subjects discriminate the different colour stimuli on the basis of their responses. Thus, we can assume that for each colour an object looks to have this science can refer to a property that is causally responsible for the fact that the stimulus looks a certain colour to Mary. Therefore, we can think that the *quale* Mary ascribes to her colour experience is a certain thick property of the nervous system that are responsible for discriminating the colours that objects look to have.

It might be objected that this account of *qualia* is inconsistent with the assumption that Mary learns something on her release. In fact, before her release, Mary knows all the causal roles that a certain property of the nervous system play. So, when she sees a coloured object she should be able to recognise the *quale* of her experience. But this criticism fails to acknowledge that by seeing colour objects Mary acquires a different way of thinking about her *qualia*.

³⁶ This is characterisation of *qualia* that has been promoted by functionalists and other philosophers, see Armstrong 1981, Lewis 1972.

³⁷ See, section 3.3, at p. 63.

When Mary sees colours she can refer to *qualia* in virtue of her new way of thinking about the colours that objects look to have. Before her release she can observe other people's discriminatory responses to certain stimuli. However, she is not directly aware of the features that they discriminate. She knows that when lights of a certain type stimulate the visual system of these subjects, they have certain discriminatory responses. However, she is not directly aware of the ways in which these stimuli look to these subjects. In her laboratory, she can only infer these features by means of certain statistical procedures that determine the colour space. Accordingly, she refers to colours as determinate positions in the colour space. On the other hand, when she is released, she becomes directly aware of the colour objects look to have. Therefore, she can formulate thoughts about the type of colour that object looks to have in virtue of being directly aware of these features. Thus, if she faces a red object she can think about the *qualia* of her red-type experience as the property that is causally responsible for the fact that the object looks red to her.

To sum up, I have suggested that a *quale* is a property of an experience that determines its causal powers. In particular, this is the power to produce certain discriminatory responses when certain stimuli conditions are given. These properties are referred to by the scientific concepts Mary possesses before her release.

7.6 Conclusion

In the present research I have offered reasons for resisting the strategy involved in Frank Jackson's knowledge argument. This is an extremely influential argument in support of the claim that conscious experiences have properties that are beyond the scope of scientific knowledge.

My conclusion is based on four main distinctive results. Firstly, I have offered a plausible formulation of the claim that science can accommodate colour experiences. By taking into account recent discussions within philosophy of science

and philosophy of mind, I have formulated this claim as the hypothesis of modest reductionism.

Secondly, I have shown that the strategy involved in Jackson's version of the knowledge argument can be used to target the modest reductionism hypothesis. Many commentators have focussed their attention on what Mary supposedly comes to know by seeing colours. Instead, I have shown that we have to take seriously the challenge, posed by Daniel Dennett and Patricia Churchland, of specifying what Mary's scientific knowledge might be. I have answered to this objection by offering an account of this knowledge in terms of contemporary psychophysics and neuroscience. The resulting formulation of the knowledge argument is weaker than Jackson's. However, it appears to threaten the modest reductionism hypothesis.

Thirdly, I have argued that the plausibility that knowledge argument must rely on an account of introspective knowledge. In particular, this account should explain the transition from Mary's seeing a coloured object to her holding a belief about the type of colour experience that she is having. I have maintained that the *Displaced Perception Model* of introspection, elaborated from an account proposed by Fred Dretske, offers this explanation. According to this model, Mary's capacities to self ascribe types of colour experiences, in virtue of seeing colours, requires certain capacities. First, she has to master colour concepts and the notion of looking a certain colour to her. Second, she has to know the systematic relations holding between things looking a certain colour to her and having a certain type of colour experience.

Finally, I have motivated a version of the two ways of thinking reply to the knowledge argument. As many opponents and supporters have recently started to realise, this strategy might be charged with being *ad hoc*. I offered a distinctive justification of this reply to the knowledge argument. Assuming the account of introspection mentioned above, the existence of visual recognitional colour concepts might justify this strategy. A person possesses these concepts when she is

able to determine the colours of objects simply by having visual experiences. On her release Mary acquires these colour concepts. Having these concepts, she can then think in new ways about the type of colour experiences she is having. However, the typology of colour experiences that she can master after her release does not require the existence of *qualia* that her scientific knowledge cannot describe or explain.

References

- Alter, T. 1995. *Knowing What it is Like*. Los Angeles: University of California, PhD Thesis.
- Alter, T. 1999. "Knowledge Argument." Web page, [accessed 31/7/2003]. Available at: <http://host.uniroma3.it/progetti/kant/field/ka.html> .
- Armstrong, D. M. 1981. "The Causal Theory of the Mind." In D. M. Armstrong. *The Nature of Mind and Other Essays*. New York: Cornell University Press. Reprinted in W. Lycan, ed. *Mind and Cognition*. Oxford: Blackwell, 1990, 37-47.
- Aydede, M. 2000. "Is Introspection Inferential?" Web page, [accessed 29/3/2003]. Available at: <http://web.clas.ufl.edu/users/maydede/dpm.pdf> .
- Aydede, M. and Güzeldere, G. 2003. "Cognitive Architecture, Concepts, and Introspection: An Information-Theoretic Account of Phenomenal Consciousness." Web page, [accessed 15/11/2003]. Available at: <http://web.clas.ufl.edu/users/maydede/introspection.pdf> .
- Barlow, H. B., Kohn, H. I., and Walsh, E. S. 1947. "Visual Sensations Aroused by Magnetic Fields." *Amer. J. Psychology* 148: 372-375.
- Bickle, J. 1998. *Psychoneural Reduction: The New Wave*. Cambridge (Mass.): MIT Press.
- Bickle, J. "Concepts of Intertheoretic Reduction in Contemporary Philosophy of Mind." Web page, [accessed 20/6/2003]. Available at: <http://host.uniroma3.it/progetti/kant/field/cir.htm> .
- Bigelow, J. and Pargetter, R. 1990. "Acquaintance with Qualia." *Theoria* 56: 129-147.
- Block, N. 1978. "Troubles with Functionalism." Reprinted (excerpt) in W. Lycan, ed. *Mind and Cognition*. Oxford: Blackwell, 1990, 444-468.
- Block, N. 1990. "Inverted Earth." *Philosophical Perspectives* 4: 53-79.
- Block, N. 1996. "Mental Paint and Mental Latex." In E. Villeneuve, ed. *Philosophical Issues*, Vol. 7. Northridge (Ca): Ridgeview.

- Byrne, A. and Hilbert, D. 2003. "Color Realism and Color Science." *Behavioral and Brain Sciences* 26: 3-64.
- Carnap, R. 1967. *The Logical Structure of the World*. Berkeley: University of California Press.
- Carruthers, P. 2000. *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.
- Carruthers, P. 2002. "Recognitional Concepts and Higher-Order Experiences." Web page, [accessed 23/7/2003]. Available at: <http://www.philosophy.umd.edu/people/faculty/pcarruthers/Phenomenal-concepts.htm> .
- Chalmers, D. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York and London: Oxford University Press.
- Chalmers, D. 1999. "Contemporary Philosophy of Mind: An Annotated Bibliography." Web page, [accessed 20/2/2001]. Available at: <http://www.u.arizona.edu/~chalmers/biblio.html> .
- Chalmers, D. 2003. "Insentience, Indexicality and Intensions." Web page, [accessed 23/7/2003]. Available at: <http://www.u.arizona.edu/~chalmers/papers/perry.html> .
- Chomsky, N. 1988. *Language and Problems of Knowledge*. Cambridge (Mass.): MIT Press.
- Churchland, P. S. 1986. *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge (Mass.): MIT Press.
- Churchland, P. 1979. *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.
- Churchland, P. 1985. "Reduction Qualia and the Direct Introspection of Brain States." *Journal of Philosophy* 82: 8-28.
- Churchland, P. 1989. "Knowing Qualia: A Reply to Jackson." In P. Churchland. *A Neurocomputational Perspective*. Cambridge (Mass.): MIT Press, 67-76. Reprinted in P. Churchland, and P. S. Churchland, *On the Contrary: Critical Essays, 1987-1997*. Cambridge (Mass.) and London: MIT Press, 1997, 143-157.
- Clark, A. 1993. *Sensory Qualities*. Oxford: Clarendon Press.

- Clark, A. 2000. *A Theory of Sentience*. Oxford: Oxford University Press.
- Conee, E. 1994. "Phenomenal Knowledge." *Australasian Journal of Philosophy* 72: 136-50.
- Coren, S., Ward, L. M., and Enns, J. T. 1999. *Sensation and Perception*. Fifth Edition. London: Harcourt Brace College Publishers.
- Crane, T. and Mellor, D. H. 1990. "There is No Question of Physicalism." *Mind* 99: 185-206.
- Crane, T. 2001. *The Elements of Mind: An Introduction to the Philosophy of Mind*. Oxford: Oxford University Press.
- Davidson, D. 1970. "Mental Events." In L. Foster and J. W. Swanson, eds. *Experience and Theory*. Amherst (Mass.): University of Massachusetts Press, 79-91. Reprinted in D. Davidson, *Essays on Actions and Events*. Second Edition. Oxford: Oxford University Press 2001, 207-225.
- De Valois, R. L. and De Valois, K. K. 1975. "Neural Coding of Color." In E. C. Carterette, and M. P. Friedman, eds. *Handbook of Perception, Vol. 5: Seeing*. Academic Press, 117-66. Reprinted in A. Byrne and D. R. Hilbert, eds. *Readings on Color. Volume Two: The Science of Color*. Cambridge (Mass.)/London: The MIT Press, 1997, 93-140.
- Dennett, D. 1988. "Quining Qualia." In A. Marcel and E. Bisiach, eds. *Consciousness in Contemporary Science*. Oxford: Oxford University Press, 43-77. Reprinted in N. Block, O. Flanagan and G. Güzeldere, eds. *The Nature of Consciousness*. Cambridge (Mass.): MIT Press, 1997, 619-642.
- Dennett, D. 1991. *Consciousness Explained*. London: Little & Brown. Reprinted London: Penguin, 1993.
- Dretske, F. 1995. *Naturalizing the Mind*. Cambridge (Mass.): MIT Press.
- Dretske, F. 1999. "The Mind Awareness of Itself." *Philosophical Studies* 95: 103-124.
- Ekman, G. 1954. "Dimensions of Color Vision." *Journal of Psychology* 38: 467-474.
- Enç, B. 1983. "In Defense of Identity Theory." *Journal of Philosophy* 80: 279-298.

- Evans, G. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.
- Feigl, H. 1934. "Logical Analysis of the Psychophysical Problem: A Contribution of the New Positivism." *Philosophy of Science* 1: 420-445.
- Feigl, H. 1960. "Mind-Body, not a Pseudo-Problem." In S. Hook, ed. *Dimensions of Mind*. New York: New York University Press. Reprinted in C.V. Borst, ed. *The Mind/Brain Identity Theory*. London: Macmillan, 1970, 33-41.
- Feigl, H. 1967. "The 'Mental' and the 'Physical' ." In H. Feigl, M. Scriven and G. Maxwell, eds. *Minnesota Studies in the Philosophy of Science. Vol. 2: Concepts, Theories and the Mind-Body Problem*. Minneapolis: University of Minnesota Press, 370-497.
- Feinberg, G. 1966. "Physics and the Thales Problem." *Journal of Philosophy* 63: 5-17.
- Feyerabend, P. K. 1962. "Explanation, Reduction and Empiricism." In H. Feigl and G. Maxwell, eds. *Minnesota Studies in the Philosophy of Science. Vol. 3: Scientific Explanation, Space, and Time*. Minneapolis: University of Minnesota Press, 28-97.
- Fodor, J. 1974. "Special Sciences (or The Disunity of Science as a Working Hypothesis)." *Synthese* 28: 97-115. Reprinted in N. Block, ed. *Readings in Philosophy of Psychology*. Vol. 1. Cambridge (Mass.): Harvard University Press, 1980, 120-133.
- Fogelin, R. J. 1984. "Hume and the Missing Shade of Blue." *Philosophy and Phenomenological Research* 45: 263-271.
- Goodman, N. 1977. *The Structure of Appearance*. Dordrecht/Boston: Reidel.
- Hardin, C. L. 1988. *Color for Philosophers. Unweaving the Rainbow*. Indianapolis and Cambridge: Hackett Publishing Company.
- Harman, G. 1990. "The Intrinsic Quality of Experience." In J. Tomberlin, ed. *Philosophical Perspectives*, Vol. 4. Atascadero (CA): Ridgeview. Reprinted in N. Block, O. Flanagan and G. Güzeldere, eds. *The Nature of Consciousness*. Cambridge (Mass.): MIT Press, 1997, 663-675.
- Hellman, G. 1985. "Determination and Logical Truth." *The Journal of Philosophy* 82: 607-16.

- Hempel, C. G. 1980. "Comments on Goodman's Ways of Worldmaking." *Synthese* 45: 139-99.
- Hooker, C. A. 1981. "Towards a General Theory of Reduction. Part I: Historical and Scientific Setting. Part II: Identity in Reduction. Part III: Cross-Categorical Reduction." *Dialogue* 20: 38-59, 201-236, 496-529.
- Horgan, T. 1984. "Jackson on Physical Information and Qualia." *Philosophical Quarterly* 34: 147-52.
- Hornsby, J. 1984. "On Functionalism, and on Jackson, Pargetter and Prior on Functionalism." *Philosophical Studies* 46: 75-95.
- Hume, D. 1978. *A Treatise of Human Nature*. Edited by L. A. Selby-Bigge and revised by P. H. Nidditch. Oxford: Clarendon Press. Reprinted
- Hurvich, L. 1981. *Color Vision*. Sunderland (Mass.): Sinauer Associates Inc.
- Ismael, J. 1999. "Science and the Phenomenal." *Philosophy of Science* 66: 351-369.
- Jackson, F. 1977. *Perception*. Cambridge: Cambridge University Press.
- Jackson, F. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly* 32: 127-36. Reprinted in W. Lycan, ed. *Mind and Cognition*. Oxford: Blackwell, 1990, 469-77.
- Jackson, F. 1986. "What Mary Didn't Know." *Journal of Philosophy* 83: 291-5. Reprinted in N. Block, O. Flanagan and G. Güzeldere, eds. *The Nature of Consciousness*. Cambridge (Mass.): MIT Press, 1997, 567-570.
- Jackson, F. 1998a. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Clarendon Press.
- Jackson, F. 1998b. "Postscript on Qualia." In F. Jackson. *Mind Method and Conditionals: Selected Essays*. London: Routledge, 76-79.
- Jackson, F., Pargetter, R., and Prior, E. W. 1982. "Functionalism and Type-Type Identity Theories." *Philosophical studies* 42: 209-225.
- Jacquette, D. 1995. "The Blue Banana Trick: Dennett on Jackson's Color Scientist." *Theoria* 61: 216-230.

- Kim, J. 1992. "Multiple Realization and the Metaphysics of Reduction." *Philosophy and Phenomenological Research* 52: 1-26. Reprinted in J. Kim. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press, 1993, 309-337.
- Kirkham, R. L. 1992. *Theories of Truth: A Critical Introduction*. Cambridge (Mass.): MIT Press.
- Kripke, S. 1972. "Naming and Necessity." In D. Davidson and G. Harman, eds. *Semantics of Natural Languages*. Dordrecht: Reidel, 253-355. Reprinted in S. Kripke. *Naming and Necessity*. Cambridge (Mass.): Harvard University Press, 1980.
- Levin, J. 1985. "Could Love Be Like a Heat-Wave? Physicalism and the Subjective Character of Experience." *Philosophical Studies* 49: 245-261.
- Levine, J. 1986. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64: 354-361.
- Levine, J. 2001. *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.
- Lewis, D. 1966. "An Argument for the Identity Theory." *Journal of Philosophy* 63: 17-25. Reprinted in D. Lewis. *Philosophical Papers*, Vol. 1. Oxford: Oxford University Press, 1983, 99-107.
- Lewis, D. 1972. "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy* 50: 291-315. Reprinted in N. Block ed. *Readings in the Philosophy of Psychology*. Vol. 1. Cambridge: Harvard University Press, 1980, 207-15.
- Lewis, D. 1979. "Attitudes *De Dicto* and *De Se*." *Philosophical Review* 88: 513-43.
- Lewis, D. 1983. "Postscript to 'Mad Pain and Martian Pain'." In D. Lewis. *Philosophical Papers*. Vol. 1. New York: Oxford University Press, 130-32.
- Lewis, D. 1986. "Causal Explanation." In D. Lewis. *Philosophical Papers*. Vol. II. Oxford: Oxford University Press, 214-40.
- Lewis, D. 1990. "What Experience Teaches." In W. Lycan, ed. *Mind and Cognition*. Oxford: Blackwell, 499-519. Reprinted in N. Block, O. Flanagan and G. Güzeldere, eds. *The Nature of Consciousness*. Cambridge (Mass.): MIT Press, 1997, 580-595.

- Loar, B. 1990. "Phenomenal States." In J. Tomberlin, ed. *Philosophical Perspective*, Vol. 4. Atascadero: Ridgeview, 81-108. Reprinted in N. Block, and O. Flanagan and G. Güzeldere, eds. *The Nature of Consciousness*. Cambridge (Mass.): MIT Press, 1997, 597-616.
- Loux, M. J. 1998. *Metaphysics*. London: Routledge.
- Lycan, W. 1995. "A Limited Defence of Phenomenal Information." In T. Metzinger, ed. *Conscious Experience*. Thorverton: Imprint Academic/Scöningh, 243-258.
- Lycan, W. 2003. "Dretske's Ways of Introspecting." Web page, [accessed 29/7/2003]. Available at:
<http://www.unc.edu/~ujanel/DretWays.htm> .
- Lyons, W. 1988. *The Disappearance of Introspection*. Cambridge (Mass.): MIT Press.
- Martin, M. G. F. 1998. "Setting Things Before the Mind." In A. O'Hear, ed. *Current Issues in Philosophy of Mind*. Cambridge: Cambridge University Press, 157-179.
- McGinn, C. 1983. *The Subjective View*. Oxford: Oxford University Press.
- McGinn, C. 1991. *The Problem of Consciousness*. Oxford: Blackwell.
- McMullen, C. 1985. "Knowing What it's Like and the Essential Indexical." *Philosophical Studies* 48: 211-233.
- Mellor, D. H. 1993. "Nothing Like Experience." *Proceedings of the Aristotelian Society* 93: 1-16.
- Melnyk, A. 1997. "How To Keep The 'Physical' in Physicalism." *Journal of Philosophy* 94: 622-637.
- Millar, A. 1991a. "Concepts, Experience and Inference." *Mind* 100: 495-505.
- Millar, A. 1991b. *Reasons and Experience*. Oxford: Clarendon Press.
- Millikan, R. G. 1990. "The Myth of the Essential Indexical." *Nous* 24: 723-734.
- Montero, B. 1999. "The Body Problem." *Nous* 33: 183-200.

- Moore, G. E. 1903. "The Refutation of Idealism." *Mind* 7: 1-30. Reprinted in G. Moore. *Philosophical Studies*. London: Routledge & Kegan Paul, 1922, 1-30.
- Moore, G. E. 1953. "Sense-Data." In G. E., Moore. *Some Main Problems of Philosophy*. London: George Allen & Unwin, 28-40. Reprinted in T. Baldwin ed. *G. E. Moore, Selected Writings*. London and New York: Routledge, 1993, 45-58.
- Nagel, E. 1961. *The Structure of Science: Problems in the Logic of Scientific Explanation*. London: Routledge & Kegan Paul.
- Nagel, T. 1974. "What is it Like to be a Bat?" *Philosophical Review* 83: 435-450. Reprinted in T. Nagel. *Mortal Questions*. Cambridge: Cambridge University Press, 1979, 165-180.
- Nagel, T. 1986. *The View from Nowhere*. New York: Oxford University Press.
- Nemirow, L. 1980. "Review of T. Nagel, *Mortal Questions*." *Philosophical Review* 89: 475-76.
- Nemirow, L. 1990. "Physicalism and the Cognitive Role of Acquaintance." In W. Lycan, ed. *Mind and Cognition*. Oxford: Blackwell, 469-77.
- Nida-Rümelin, M. 1995. "What Mary Couldn't Know." In T. Metzinger, ed. *Conscious Experience*. Thorverton: Schöningh/Academic Press, 219-241.
- Oppenheim, P. and Putnam, H. 1958. "Unity of Science as a Working Hypothesis." In H. Feigl, M. Scriven, and G. Maxwell, eds. *Minnesota Studies in the Philosophy of Science*. Vol. 2. Minneapolis: University of Minnesota Press, 3-36. Reprinted in R. Boyd, P. Gasper and J. D. Trout, eds. *The Philosophy of Science*. Cambridge (Mass.): MIT Press, 405-427.
- Palmer, S. E. 1999a. "Color, Consciousness and the Isomorphism Constraint." *Behavioral and Brain Sciences* 22: 923-989.
- Palmer, S. E. 1999b. *Vision Science*. Cambridge (Mass.): MIT Press.
- Papineau, D. 1993. *Philosophical Naturalism*. Oxford: Basil Blackwell.
- Papineau, D. 1994. "Content (2)." In S. Guttenplan, ed. *A Companion to the Philosophy of Mind*. Oxford: Blackwell, 225-230.
- Papineau, D. 2002. *Thinking about Consciousness*. Oxford: Clarendon Press.

- Peacocke, C. 1983. *Sense and Content: Experience, Thoughts, and their Relations*. Oxford: Clarendon Press.
- Peacocke, C. 1984. "Colour Concepts and Colour Experience." *Synthese* 58: 365-82.
- Peacocke, C. 1989. "Perceptual Content." In J. Almong, J. Perry H. Wettstein, eds. *Themes from Kaplan*. New York and Oxford: Oxford University Press, 297-329.
- Peacocke, C. 1992. *A Study of Concepts*. Cambridge (Mass.): MIT Press.
- Peacocke, C. 1994. "Content (1)." In Samuel Guttenplan, ed. *A Companion to the Philosophy of Mind*. Oxford: Blackwell, 219-225.
- Penfield, W. 1958. *The Excitable Cortex in Conscious Man*. Liverpool: Liverpool University Press.
- Perry, J. 1977. "Frege on Demonstratives." *Philosophical Review* 86: 474-497.
- Perry, J. 2001. *Knowledge, Possibility and Consciousness: The 1999 Jean Nicod Lectures*. Cambridge (Mass.): MIT Press.
- Pitcher, G. 1964. "Introduction." In G. Pitcher, ed. *Truth*. Englewood Cliffs (NJ): Prentice-Hall, 1-15.
- Place, U. T. 1956. "Is Consciousness a Brain Process?" *British Journal of Psychology* 47: 243-55.
- Poland, J. 1994. *Physicalism: The Philosophical Foundations*. Oxford: Clarendon Press.
- Prior, E., Pargetter, R., and Jackson, F. 1982. "Three Theses about Dispositions." *American Philosophical Quarterly* 19: 251-7.
- Prior, E. 1985. *Dispositions*. Aberdeen: Aberdeen University Press.
- Putnam, H. 1967. "Psychological Predicates." In W. H. Capitan and D.D. Merrill, eds. *Art, Mind and Religion*. Pittsburgh: University of Pittsburgh Press. Reprinted in as "The Nature of Mental States" in H. Putnam. *Mind, Language, and Reality. Philosophical Papers*, Vol. 2. Cambridge: Cambridge University Press, 1975, 429-440.

- Raymont, P. 1999. "The Know-How Response to Jackson's Knowledge Argument." *Journal of Philosophical Research* 24: 113-126.
- Rey, G. 1992. "Sensational Sentences." In M. Davies and G. Humphreys, eds. *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.
- Robinson, H. 1993. "Dennett on the Knowledge Argument." *Analysis* 53: 174-77.
- Robinson, H. 1994. *Perception*. London: Routledge.
- Russell, B. 1912. *The Problems of Philosophy*. London: Williams & Norgate. Reprinted London: Oxford University Press, 1967.
- Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson. Reprinted with an introduction by D. Dennett, Penguin, London, 2000.
- Salmon, W. C. 1990. *Four Decades of Scientific Explanations*. Minneapolis: University of Minnesota Press.
- Schaffner, K. 1967. "Approaches to Reduction." *Philosophy of Science*, 34: 137-147.
- Seager, W. 1991. *Metaphysics of Consciousness*. New York: Routledge.
- Seager, W. 1999. *Theories of Consciousness: An Introduction and Assessment*. London and New York: Routledge .
- Sellars, W. 1963. *Science, Perception, and Reality*. London: Routledge & Kegan Paul.
- Shepard, R. N. 1962a. "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function: Part I." *Psychometrika* 27, 3: 125-40.
- Shifman, S. S., Reynolds, M. L., and Young, F. W. 1981. *Introduction to Multidimensional Scaling*. New York: Academic Press.
- Shoemaker, S. 1980. "Causality and Properties." In P. van Inwagen, ed. *Time and Cause*. Dordrecht: D. Reidel, 109-35. Reprinted in D. H. Mellor and A. Oliver, eds. *Properties*. Oxford: Oxford University Press, 1997, 228-254.

- Shoemaker, S. 1996. "Self-knowledge and "inner sense". Lecture I: The object perception model." In S. Shoemaker. *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press, 201-223.
- Smart, J. J. C. 1959. "Sensations and Brain Processes.": Reprinted (revised version) in C. V. Borst ed. *The Mind/Brain Identity Theory*. London: Macmillan, 1970, 52-66.
- Smart, J. J. C. 1978. "The Content of Physicalism." *The Philosophical Quarterly* 28: 239-41.
- Smart, J. J. C. 1989. *Our Place in the Universe*. Oxford: Oxford University Press.
- Smith, P. 1992. "Modest Reduction and the Unity of Science." In D. Charles and K. Lennon, eds. *Reduction, Explanation and Realism*. Oxford: Clarendon Press, 19-43.
- Snowdon, P. 1992. "How to Interpret 'Direct Perception'." In T. Crane, ed. *The Contents of Experience*. Cambridge: Cambridge University Press, 48-104.
- Stoljar, D. 2001. "Physicalism." in *Stanford Encyclopaedia of Philosophy*, Web page, [accessed 27/6/2001]. Available at: <http://cd1.library.usyd.edu.au/stanford/entries/physicalism/> .
- Strawson, G. 1989. "Red and 'Red'." *Synthese* 78: 193-232.
- Sturgeon, S. 2000. *Matters of Mind. Consciousness Reason and Nature*. London and New York: Routledge.
- Teller, D. Y. 1984. "Linking Propositions." *Vision Research* 24, 10: 1233-1246.
- Trout, J. D. 1991. "Reductionism and the Unity of Science: Introductory Essay." In R. Boyd, P. Gasper and J. D. Trout, eds. *The Philosophy of Science*. Cambridge (Mass.): MIT Press, 387-392.
- Tye, M. 1983. "Functional and Type-Physicalism." *Philosophical Studies* 44: 161-174.
- Tye, M. 1984. "The Adverbial Approach to Visual Experience." *Philosophical Review* 93: 195-225.
- Tye, M. 1992. "Visual Qualia and Visual Content." In T. Crane, ed. *The Contents of Experience*. Cambridge: Cambridge University Press.

Tye, M. 1995. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge (Mass.): MIT Press.

Tye, M. 2000. *Consciousness, Color and Content*. Cambridge (Mass.) and London: MIT Press.

Tye, M. 2003. "A Theory of Phenomenal Concepts." Web page, [accessed 9/10/2003]. Available at:
<http://www.utexas.edu/cola/depts/philosophy/faculty/tye/Theory.pdf> .