# The Knowledge Argument and Phenomenal Concepts

# The Knowledge Argument and Phenomenal Concepts

By

Luca Malatesti

CAMBRIDGE SCHOLARS

PUBLISHING

The Knowledge Argument and Phenomenal Concepts,
by Luca Malatesti

This book first published 2012

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library



ISBN (10): 1-4438-4060-2, ISBN (13): 978-1-4438-4060-6

To Nela and Antonio
for all the love and fun

# TABLE OF CONTENTS

# ACKNOWLEDGEMENTS

Much of this book originates from my doctoral dissertation at the University of Stirling. I would like to thank José Luis Bermúdez and Alan Millar for their supervision and encouragement.

In recent years I have tried out my ideas in a number of places, including the Departments of Philosophy at the Universities of Arezzo, Birmingham, Hull, Stirling, Firenze, Pécs and Rijeka. I would like to express my gratitude to those who invited me and to the students, colleagues and friends who engaged me in challenging discussions.

I thank the relevant publishers and editors for permission to reproduce parts or revisions of my published work. Chapter 3 contains materials from my paper "Mary's Scientific Knowledge," *Prolegomena* 7 (2008): 37–59. Chapter 6 contains a substantial part of the chapter "Knowing What it is Like and Knowing How," in *Mind and Causality*, ed. A. Peruzzi (Amsterdam: John Benjamins, 2004), 119–129. Chapters 7 and 8 contain parts, with modifications, of my paper "Phenomenal Ways of Thinking," *Teorema* 27, no. 3 (2008): 149–166.

Finally, I thank Ben Young, of Babel Editing, for his fast and accurate copy-editing.



The printing of this book was partly financed by the Foundation of the University of Rijeka (Zaklada Sveučilišta u Rijeci), under the contract class: 612-10/10-01/54, reference number: 2170-57-06-10-2. The opinions expressed in this book are of the author and do not represent those of the Foundation of the University of Rijeka.

# INTRODUCTION

## 1. Consciousness in the natural world

There is widespread debate in contemporary philosophy of mind over the place of conscious experiences in the natural world – where the latter is taken to be broadly as described and explained by such sciences as physics, chemistry and biology.[1] Conscious experiences encompass pains, bodily sensations, perceptions, feelings and moods. Many philosophers maintain that these mental states can be completely described and explained in scientific terms. This claim is usually seen as related to the ontological statement that mental states are or depend on (in ways to be specified) physical states of the brain. Others argue, however, that conscious experiences pose a fundamental problem for scientific knowledge. Authors in this camp attack the claim that scientific knowledge can accommodate these mental states.[2] According to these philosophers, when we perceive a yellow lemon or endure a pain, certain properties are instantiated that can be neither described nor explained by science.

Frank Jackson advanced the *knowledge argument* to support the claim that conscious colour experiences constitute an insoluble problem for science.[3] In recent years he has recanted;[4] however, his reasoning is still very influential.[5] The argument is based on a thought experiment concerning Mary, a vision scientist who has complete scientific knowledge of the nature of colours and colour vision. However, she herself has never had a colour experience, because she has been trapped since birth in an entirely black-and-white environment. According to Jackson's knowledge

[1] For introductions to this and related philosophical issues concerning consciousness, see Block, Flanagan and Güzeldere 1997 and Velmans and Schneider 2007.

[2] Influential formulations of this criticism can be found in Chalmers 1996, McGinn 1991 and Nagel 1974.

[3] Jackson 1982 and 1986.

[4] Braddon-Mitchell and Jackson (1996, 134) maintain that there are good reasons for thinking that the conclusion of the knowledge argument is false, so the argument itself must go wrong somewhere. Jackson offers his diagnosis of what is wrong in Jackson 2004 and 2007.

[5] For two excellent collections of papers on the knowledge argument, see Ludlow, Stolijar and Nagasawa 2004 and Alter and Walter 2007.

argument, upon being released and seeing coloured objects for the first time, Mary acquires new knowledge that had previously escaped her complete scientific knowledge. Thus, the argument goes, we should conclude 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*.

Qualia are taken to be properties that specify what it is like from the perspective of the subject to undergo a certain conscious experience. For example, in the case of colour experiences, qualia are often understood as the ways in which colours are given to the subject that experiences them. So, in having the conscious experience of a red rose, the subject faces a *quale* that is the way in which red looks to her. In addition, qualia are taken to be proprieties that categorise conscious experiences. For instance, pains differ from experiences of colours because they have different qualia. Similarly, the difference between a conscious experience of red and one of blue is based on a difference between their qualia.

This book 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. I argue in Chapter 1 that the hypothesis that science can account for colour experiences should be formulated as the *modest reductionism hypothesis*. Roughly, this is the hypothesis that a certain scientific theory, which can be explanatorily interfaced with our current physics of ordinary matter, can account for conscious experiences. In addition, we have to consider a certain *version* of the knowledge argument. As presented above, Jackson's argument involves the unintelligible premise that Mary has complete (future or possible) scientific knowledge.[6] Without reformulating this premise in a way that renders it intelligible, 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 2, I address the question of Mary's scientific knowledge. First I explain how the preliminary characterisation of her knowledge is intelligible, and then show how it may be helpfully elaborated on the basis of the descriptions and explanations of colour experiences that are involved in current physics and neuroscience. I then argue that modest reductionists can plausibly assume that the scientific knowledge delineated by this characterisation can account for colour experiences.

---

[6] This criticism is advanced in Churchland 1986, 331–334. Daniel Dennett advances a similar line of criticism in Dennett 1991, 399–403, and it is developed in more detail in Dennett 2007.

The main thesis of this book is that the plausible version of the knowledge argument does not threaten modest reductionism. This conclusion is based on the endorsement of what can be called the "phenomenal concept strategy."[7] According to this response, what the knowledge argument shows is that there are different ways of thinking about colour experiences. One way is provided by scientific knowledge; another is offered by our ordinary conception of colour experiences. However, the knowledge argument fails to establish its ontological conclusion: the 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 on the basis of two investigations. The first line of research, carried forward in Chapters 2–4, seeks to reveal and evaluate the implicit assumptions that figure in the knowledge argument. The second main investigation, which begins in Chapter 5 and continues in Chapters 6 and 7, assesses the knowledge argument itself. Specifically, in these chapters I justify my adoption of the *phenomenal concept strategy*. The remainder of this introduction sketches these two investigations and so offers an outline of the book overall.

## 2. The implicit premises of the knowledge argument

A considerable part of this volume is dedicated to the elucidation of the implicit assumptions that figure in the knowledge argument. The vast literature on this argument contains important insights about its structure and the meaning of its premises.[8] By considering and evaluating these different suggestions, I offer a comprehensive analysis of the knowledge argument.

The elucidation of the knowledge argument will lead us to an important result: namely, that 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.

According to a standard interpretation, Mary comes to know that experiences have qualia. This interpretation requires that, by seeing coloured

[7] This reply is also called the "new mode of presentation reply," the "conceptual dualism strategy" and the "new knowledge and old fact reply." The strategy is upheld, for example, in Peacocke 1989, Loar 1990, Papineau 2002, Sturgeon 2000, Carruthers 2000, Tye 2000 and Perry 2001. Precursors of the kind of view involved in this reply can be found in Feigl 1960, Smart 1959 and Place 1956.

[8] Detailed analyses of the structure and implicit premises of the knowledge argument can be found in Alter 1995, Perry 2001, Lewis 1990 and Churchland 1989.

objects, Mary can form beliefs about the type of colour experience that 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 these new beliefs that colour experiences have qualia. Such an explanation requires accounting for the transition from seeing coloured objects to acquiring these beliefs. Chapters 3 and 4 consider how the upholder of the knowledge argument might offer an account of this. 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

Chapters 5, 6 and 7 consider whether Mary, by acquiring new knowledge about colour experiences, comes to know facts that escape her scientific knowledge. I will endorse a reply to the knowledge argument that has come to be, in effect, a standard move amongst many philosophers. This reply, which has been called the *phenomenal concept strategy*, assumes the consistency of the following claims. First, that upon her release Mary acquires new beliefs about her colour experiences. Second, that 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 have possessed before her release. These concepts enable her to have new thoughts and thus new beliefs about colour experiences.

My account of the phenomenal concept strategy is distinctive in two respects. First, 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 *ad hoc*.[9] 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.[10]

The second distinctive element of my formulation of the phenomenal concept strategy derives from the view on introspection which I defend. I argue that this strategy should be grounded on the existence of *recognitional concepts* of colours. Having colour experiences is a requirement for possessing and applying these concepts. 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 a description of Jackson's knowledge argument. After analysing this argument, I show that it is based on the unintelligible premise that Mary possesses complete scientific knowledge of colour and colour vision. However, what emerges from this discussion is that the general strategy involved in this argument might still be viable as a way to target the modest reductionism hypothesis. In particular, by considering contemporary psychophysics and neuroscience, I offer an intelligible and plausible characterisation of Mary's scientific knowledge. Specifically, this characterisation satisfies the general requirement of the modest reductionism hypothesis. In addition, this account of Mary's scientific knowledge can be used in a version of the knowledge argument.

In Chapter 3 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 stands in need of 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

[9] For this criticism, see Levine 2001, 86.

[10] Recent attempts to justify the phenomenal concept strategy can be found in Perry 2001, Papineau 2002, Carruthers 2000, Tye 2003 and Aydede and Güzeldere 2005.

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 4 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 the introspective knowledge of colour experiences. From this account, I derive a characterisation of the beliefs that figure in Mary's supposedly new knowledge. 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 5 evaluates 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 regarding the individuation of beliefs. Therefore, we can concede to such a supporter that Mary acquires new propositional knowledge about her colour experiences.

Chapter 6 begins with an illustration of the phenomenal concept reply to the knowledge argument. Following this, I criticise a version of this strategy elaborated by John Perry.[11] The central thesis of this account is that the phenomenal concepts that Mary acquires by having colour experiences are particular demonstrative concepts. I will argue that Perry's account of the ways in which Mary can come to think about colour experiences once she sees colours fails to provide a satisfactory analysis of phenomenal concepts.

In Chapter 7 I show that the *quotational account* of phenomenal concepts offered by David Papineau is unsatisfactory.[12] The central idea in this approach is that phenomenal concepts are partly constituted by the experiences they refer to. I will argue that although Papineau's theory of phenomenal concepts solves the problems encountered by Perry's account, it faces other difficulties. The principal problem is that the quotational account cannot be formulated in terms consistent with physicalism. Finally, I will advance a version of the phenomenal concept strategy against the

[11] Perry 2001.
[12] Papineau 2002, 2007.

knowledge argument that is based on recognitional concepts concerning colours.

# CHAPTER ONE

# PHYSICALISM IN THE PHILOSOPHY OF MIND

## 1. Introduction

This chapter sets out a preliminary discussion of physicalism in the philosophy of mind, preparatory to the investigation of Frank Jackson's knowledge argument as embarked upon in the next chapter. Physicalists have offered an influential version of the hypothesis that our mental life in general and our conscious experiences in particular can be described and explained in scientific terms. Physicalists maintain that physics should have a fundamental role in the study of mind. They assume that physics will eventually offer a unified science that will provide an exhaustive account of the mind.

Section 2 introduces physicalism and considers some difficulties in the characterisation of the notion of the physical. Section 3 offers what I call *modest physicalism* as an account of the physical which is aimed at avoiding these problems. However, even with a notion of the physical in place, in order to formulate the requisite notion of physicalism one needs to explain in some detail how one expects to accommodate conscious experiences within the natural world. Sections 4 and 5 illustrate the debate on this issue as played out between reductive physicalists and non-reductive physicalists. Section 6 will show that, although reductionism is untenable, physicalists can still promote a plausible programme for the unification of the study of the mind and physical science. Such a programme, here called *modest reductionism*, is formulated in section 7.

## 2. Defining the physical

A central hypothesis has shaped contemporary physicalism. Following David Lewis, we can call it the hypothesis of the *explanatory adequacy of physics*.[1] This claim is:

[1] For an account of this principle, see Papineau 2002, 232–256.

> 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). (Lewis 1966, 23)

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)

The central idea in this passage is that 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 any scientific theory which differs from theoretical physics, is constituted by entities that ultimately are composed of the basic entities posited by physics.

The hypothesis of the explanatory adequacy of physics entails that if a physical phenomenon is explained by means of another phenomenon, then this latter phenomenon has to be physical.[2] In particular, on 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.[3] 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; for this, 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 proposal by referring to Lewis. He argues that mental states cause physical effects and, thus, that they figure in the explanation

[2] Lewis 1966, 23–24.

[3] Some physicalists directly assume the thesis of the causal closure of the physical, avoiding any reference to explanatory adequacy. See, for instance, Papineau 2002, 17–18.

of physical facts.[4] The hypothesis of the explanatory adequacy of physics states that only physical phenomena can explain other physical phenomena. Therefore, mental states are physical phenomena.

For example, Lewis 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 modifications in determinate muscles. Thus, pains play a role in the explanation of certain physical phenomena, and so, 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.[5]

Lewis's physicalist view delineated in the previous section involves what can be called a *theory-based conception* of being physical.[6] 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*.[7] 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 physical theory, or it is grounded in some ideally complete future (or possible) physical theory. Geoffrey Hellman describes the dilemma that emerges 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

[4] See Lewis 1966, 19–22. Lewis endorses and elaborates 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 (see, for example, Papineau 2002, 38–39) maintain that the mental is casually efficacious with respect to the physical is a simple matter of fact that does not have to be reflected in our concepts concerning mental states.

[5] Lewis 1966, 24.

[6] I take the name of this account from Stoljar 2001.

[7] Hempel 1980, see also Crane and Mellor 1990.

> "physical object, property, or law" framed independently of existing physical theories. (Hellman 1985, 609)

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. However, 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* –indeed, the historical evidence seems to point to the contrary. Physics has undergone radical theoretical revolutions in the past, and thereby come to include new entities in its ontology. For example, eighteenth-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 the theory-based conception of being physical which refers to a future (or possible) completed physics. This line of argument faces two problems. 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 present we should be agnostic about the issue.

There is a second problem with the appeal to a future completed 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 threatens to become a mere truism. We cannot exclude the possibility that a physical theory will involve reference to irreducible mental properties. Of course, stating this as a mere possibility is not in itself a problem for physicalists. They might simply claim, in response, that a complete physics might *not* make reference to mental entities. However, some philosophers argue that completing theoretical physics even as it is presently understood might actually *require* reference to conscious mental 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, and the fundamental principle of quantum mechanics is represented by the Schrödinger equation. This differential equation predicts the dynamics of the wave functions that describe the basic particles, and its underlying principle requires that the properties of basic particles, such as their position or momentum, do 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 equation. For this reason, something called the *measurement principle* is introduced; this states that when we observe particles, the wave function does not behave as predicted by the Schrödinger equation. 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 commentators have argued that the solution to the problem of the interpretation of quantum mechanics requires referring to conscious states of the observer. David Chalmers has advanced a proposal of this nature.[8] He has argued that in reality the only principle required for quantum mechanics is that expressed by the Schrödinger equation. In addition, however, he proposes that the effects described via the measurement principle can be more easily explained if we assume the existence of non-reducible conscious states. Of course, I am not here endorsing this as a solution to what is a very hard problem 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 difficulties of 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.[9] Some authors, finding Hempel's dilemma convincing, have thus sought an alternative to the theory-based conception of the physical.[10] A supporter of physicalism could, for example, avoid the preceding difficulties by endorsing an *object-based conception* of the physical, such

[8] Chalmers 1996, 333–357.

[9] See Montero 1999 and Levine 2001.

[10] For reactions to Hempel's dilemma, see Hellman 1985, Melnyk 1997, Montero 1999 and Levine 2001. See also the discussion in Poland 1994, chap. 3, 157ff.

that physical entities are introduced as being of the same type as some entity that is taken to be paradigmatically physical.[11] The central idea in this account is that the paradigmatically physical entities can be introduced without any reference to contemporary, future or possible science. Non-sentient ordinary objects such as tables and chairs can be taken to be paradigmatically physical. Thus, we might define physical properties and relations as those that are required in order to describe the paradigmatic physical entities.

The object-based conception of the physical faces problems. It is threatened, for example, by the possibility of *panpsychism*.[12] Panpsychists believe that every entity has a mind. On this view, the paradigmatically physical objects referred to in the object-based account would themselves have minds; the object-based account, however, needs to define "physical" by making originary reference to non-sentient objects. The problem here is that although the promoters of the object-based account have to exclude the possibility of panpsychism, their theory does not seem to possess the resources to further specify the nature of these non-sentient objects. It seems that they are simply excluding the possibility of panpsychism by definition, and although panpsychism might strike us as completely implausible, it 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 an instance of a deeper problem with the object-based account. This account of "physical" 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. 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 an account of ordinary objects are very different from those we might contemplate in our ordinary experience.[13] Thus, a form of physicalism based on the object-based account might well even contrast with physics itself. Endorsing such a theory therefore requires countenancing the possibility that many of the basic assumptions of contemporary physics are false. But not many physicalists would be

[11] See Jackson 1998, 6–7.
[12] Jackson 1998, 7.
[13] A criticism of this type can be found in Levine 2001, 20.

willing to accept an account of physical reality that was different from that provided by contemporary physics.

The problems that afflict the theory-based and object-based conceptions of being physical might be avoided by denying one crucial assumption that they share – namely, the idea that the formulation of physicalism does indeed stand in need of a characterisation of physical entities.[14] Joseph Levine, for example, argues that in the philosophical discussion on the mind–body problem, physical properties should be defined *per via negativa*. He argues that we have a clear enough grasp of the mental properties. In particular, the mental is characterised by *phenomenal properties*, which specify what it is like to have mental states, and *representational* properties, which determine the content of mental states. Thus, he suggests that physicalism (in his words "materialism") should be understood as the thesis that non-mental properties have ontological and explanatory priority over mental ones.[15]

This characterisation of "physical" *per via negativa* should be rejected. In fact, it undermines a very good reason for endorsing physicalism. Physicalists place confidence in the explanatory power of scientific knowledge in virtue of the actual developments of contemporary science. In particular, many physicalists have been impressed by the results of biology and neuroscience, which have explained many aspects of both normal and pathological human behaviour.[16] Thus, it is central to the physicalist project to offer a conception of the mind that is not only consistent with contemporary science, but which could also aid scientific progress.[17] Clearly, justifying physicalism by referring to current scientific practice must involve a theory-based conception of the physical. However, this leads us back to Hempel's dilemma. Therefore, we must investigate whether the physicalist can meet this difficulty without abandoning the theory-based conception of being physical.

[14] This account is offered in Montero 1999. See also Levine 2001, 20–21.
[15] Levine 2001, 21.
[16] David Papineau maintains that the idea of the 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, 232–256.
[17] See Fodor 1974, Smart 1959 and Churchland 1986.

## 3. Modest physicalism

One of the horns of Hempel's dilemma, in brief, is that if "physical" is defined by reference to future physical science, we lack any substantive grasp of the claimed 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 possible future changes.[18] However, a less demanding enterprise would be to see whether the other horn of Hempel's dilemma can be avoided.

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.[19] For we cannot exclude that future physics might consider new fundamental particles governed by laws that we do not currently know. 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 this is possible, then it might follow that the second horn of Hempel's dilemma is not as serious as first appears. Let us consider this option.

Physicalism in the philosophy of mind can be formulated as the view that mental phenomena can be explained in terms of properties precisely of the kind recognised by current physical science. J. J. C. Smart has advanced such a position by tying physicalism, understood as a theory about the mind, to contemporary physics.[20] 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 the principles and properties of current physics. 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 transformations in those theories that consider the whole universe. 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

[18] An attempt along these lines is set out in Poland 1994.

[19] Another interesting defence of a theory-based account which 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.

[20] Smart 1978, 1989.

fact that the hydrogen atom contains one proton and one electron and that water is $H_2O$.

The second tenet of Smart's position is that the contemporary physics of ordinary matter can account for the nature and workings of the mind. He assumes that "The properties of mind depend on the properties of 'ordinary matter'."[21] Specifically, he claims that the properties of the mind depend on those of the brain. Smart concedes that there are gaps in our understanding of the functioning of the brain and of the mind. However, these derive merely from difficulties in the specific application of known principles of chemistry and physics.[22] Therefore, we should not expect that "discoveries about quarks, black holes, theories of strings and superstrings" would affect our knowledge of the mind.[23]

Smart proposes that the idea that scientific knowledge can accommodate the mind is a substantive doctrine that is not obviously false. This view is arrived at by formulating physicalism properly, and 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 mental states and processes involve these macroscopic phenomena. Thus, we have a plausible formulation of the main intuitions involved in the physicalist solution to the mind–body problem. However, to be completely satisfactory this formulation needs some refinements.

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.[24] On this view, certain relations between an individual and her environment figure in the conditions that specify her mental states. However, it seems that many of these externalist authors also acknowledge that physics might be able to describe and explain these causal relations. Similarly, they appear to be willing to accept that physics is able to explain the relevant features of the external objects on which mental properties might depend. However, efforts to specify the relation of dependence between mental properties and physical properties might create another difficulty.

---

[21] Smart 1978, 341.
[22] Smart 1978, 349.
[23] Smart 1989, 80.
[24] In particular, externalist accounts of conscious experiences have been proposed in Dretske 1995 and Tye 1995.

Philosophers such as Smart and Lewis have formulated physicalism as a reductionist doctrine. Reductionism involves interrelated ontological and epistemological theses. Reductionists state that types of mental phenomena are identical to types of physical phenomena. On the epistemological side, reductionists account for reduction as a relation between scientific theories. Consequently, they claim that all 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.

## 4. The classical account of reduction

Physicalists have articulated different versions of the project of unification of the study of the mind with the study of the natural world. A variety of epistemological and ontological theses have been proposed in this respect. From the epistemological point of view, physicalists have differed with respect to their accounts of the relation between the scientific theories studying the mind and those studying the physical world. On the ontological side, they have promoted different views about the relation between mental and physical entities, consistent with their epistemological assumptions.

An influential version of contemporary physicalism has expressed the physicalist programme in the idiom of reductionism. According to reductionists, there is a trend in contemporary science which vindicates the idea that all scientific knowledge is reducible to physical knowledge.[25] Obviously, the notion of intertheoretic reduction is central to this position.

One influential view on intertheoretic reduction focuses on the explanatory capacities of scientific theories.[26] 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 $T_1$.[27] Ernest Nagel has provided a classic account of such an explanatory subsumption of theories.[28] In his formal analysis, a theory $T_2$ is reduced to a theory $T_1$ when all the statements of $T_2$ have been shown to follow from the statements of $T_1$.

[25] For a classical statement of this project, see Oppenheim and Putnam 1958.
[26] It is important to note that a different attempt has been made to characterise the idea of the unification of scientific knowledge by means of reductionism. This is the programme of semantic reduction. On this view, all theoretical sentences can be proved logically equivalent to observational sentences via the definition of theoretical terms by means of observational terms. For more on this reductive programme, see Trout 1991, 388–389.
[27] See Oppenheim and Putnam 1958.
[28] Nagel 1961, 337–397.