Metaphysical Naturalism and the Ignorance of Categorical Properties

A thesis submitted in partial fulfilment of requirements for the degree of Doctor of Philosophy

Lok-Chi Chan
The University of Sydney
Department of Philosophy
March, 2017
Copyright © 2017 by Lok-Chi Chan

Thesis supervisor: Professor David Braddon-Mitchell
Department of Philosophy, The University of Sydney
Cover design: Chan Man Yuk
Declaration

I certify that the intellectual content of this thesis is the product of my own work. Assistance received in preparing this thesis and sources have been acknowledged.

Lok-Chi Chan

Lok-Chi Chan
Abstract

The purpose of this dissertation is to investigate the connections between metaphysical naturalism and the categorical ignorance thesis – offered by Rae Langton (1998), David Lewis (2009), Frank Jackson (1998), and Simon Blackburn (1990) – and determine whether the latter will challenge the former. According to metaphysical naturalism, the actual world contains only metaphysically natural things. According to the categorical ignorance thesis, all we can know about things are their dispositional properties, but the categorical properties that bear these properties remain in principle unknowable. In this dissertation, I will determine whether the ignorance of categorical properties – as Rae Langton (1998), David Braddon-Mitchell and Frank Jackson (2007), John Foster (1993) and Alyssa Ney (2007) argue (or worry about) – is consistent with metaphysical naturalism (or physicalism), and whether it will lead to a scepticism about the metaphysical naturalness of categorical properties, which will, in turn, significantly decrease the probability that metaphysical naturalism is true. Being attracted to metaphysical naturalism, the categorical ignorance thesis, and also the philosophical position that the two theses are consistent, I will argue that the answer to the former question is yes, and that the answer to the latter question is no.

Russellian Monism, an influential doctrine in philosophy of mind that is relevant to the topic, will also be considered. According to the doctrine, our first-person experiences are constituted by their categorical bases in some manner that is not (solely) via the dispositions borne by these bases. Some philosophers, such as Bertrand Russell (1992/1927a, 1927b), thus consider categorical properties to be knowable through our acquaintance with those experiences. On the other side, some philosophers, such as Daniel Stoljar (2001a, 2001b, 2006), David Chalmers (1996), and Galen Strawson (2013), argue that Russellian Monism is compatible with metaphysical naturalism (or physicalism). In this dissertation, I will determine whether the acquaintance view put forward by some Russellian Monists, understood as an objection to the categorical ignorance thesis, is true and whether it is compatible with metaphysical naturalism; I will argue that the answers to both questions are no.

Keywords: naturalism; physicalism; Russellian Monism; categorical properties; intrinsic properties; qualia
To my most important mentors,
Belinda, David, and Graham
This page intentionally left blank
Acknowledgements

I am heavily indebted to many mentors, colleagues, friends, and my family. Firstly, I wish to thank my current and former supervisors. They are David Braddon-Mitchell, my current supervisor at the University of Sydney, Graham Oppy, my honours supervisor at Monash University, and Daniel Stoljar, my mentor during my visit to the Australian National University. I am grateful for the rich philosophical knowledge, the interesting ways of thinking and the writing skills these three great philosophers have taught me. Many of my debts to them should be apparent as this dissertation goes on. More importantly, what they have taught me is not limited to academic knowledge and skills. In particular, I will be forever grateful to David for his enthusiastic support in every respect.

I was and am very fortunate to have been working with a colleague, Andrew James Latham, on many joint research projects up until the present day. In this process, he has contributed many important ideas to this dissertation.

My parents, On-Fat Chan and Flora Chung, are of course on the list. Without their selfless and unconditional support, I could not have even started studying philosophy and working on this dissertation in the first place (let alone the fact that I would not have existed in the first place).

For helpful discussions, comments and/or non-academic assistance in the course of this research, and/or on my way towards doing it, I am grateful to Frank Jackson, John Bigelow, Jakob Hohwy, Mark Colyvan, Tristram McPherson, Toby Handfield, Karen Green, Nicholas J.J. Smith, Kristie Miller, Wai-hung Wong, Monima Chadha, Alyssa Ney, David Macarthur, Maureen O'Malley, Michael Duncan, Ben Blumson, Sam Baron, Luke Roelofs, Ryoji Sato, Simon Varey, Alex Sandgren, James Norton, Belinda Calderone, Kristall Rozen, Man Yuk Chan, Lina Ginz, Wayne Chung, Erick Llamas, Naoyuki Kajimoto, Donald Nordblom, Rose Kyo, Kevin Lam, Priscilla Ho, Hei Man Ho and finally, my secondary school teacher who seriously recommended studying philosophy to me, Ms. Au.

A special appreciation and gratefulness is due to a good friend and mentor of mine, Belinda Rickard, with whom I did not only have a good exchange of philosophical ideas, but was also educated by her much richer life experience and insight, without which I could not possibly work with the spirit and insight I had when I was writing this dissertation. This dissertation is dedicated to her, David and Graham.
This page intentionally left blank
Contents

Declaration .............................................................................................................................. ii
Abstract ................................................................................................................................. iii
Acknowledgements ............................................................................................................. vi
List of abbreviations ........................................................................................................... x
Introduction .......................................................................................................................... 1

Part I: Metaphysical naturalism ......................................................................................... 7
Section 1. A method for characterising metaphysical naturalism ........................................ 9
Section 2. Characterising metaphysical naturalism ............................................................. 23
Section 3. The content of metaphysical naturalism ............................................................ 52

Part II: Categorical ignorance ............................................................................................. 77
Section 4. Introduction to categorical properties ............................................................... 79
Section 5. The existence of categorical properties .............................................................. 87
Section 6. The nature of categorical properties ................................................................... 112
Section 7. Categorical ignorance ....................................................................................... 123

Part III: Russellian Monism ............................................................................................... 143
Section 8. Russell and Stoljar’s Russellian Monism ............................................................ 145
Section 9. The categorical acquaintance thesis ................................................................. 166
Section 10. The naturalistic unacceptability of Russellian Monism ..................................... 189
Section 11. Qualia, the explanatory gap, and metaphysical naturalism ............................... 202

Part IV: Metaphysical naturalism and categorical ignorance ............................................. 229
Section 12. Metaphysical naturalism and categorical ignorance ......................................... 231
References ........................................................................................................................... 244
Appendix A. The meaning(s) of the term ‘naturalism’ ....................................................... 259
Appendix B. The historical path approach to defining metaphysical naturalism and its rivals ........................................................................................................... 270
This page intentionally left blank
### List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN</td>
<td>Metaphysical naturalism. See p. 1.</td>
</tr>
<tr>
<td>CI</td>
<td>The categorical ignorance thesis. See p. 3.</td>
</tr>
<tr>
<td>RM</td>
<td>Russellian Monism. See p. 4.</td>
</tr>
<tr>
<td>SCI</td>
<td>The scientific categorical ignorance thesis. See p. 4.</td>
</tr>
<tr>
<td>CQ</td>
<td>The categorical qualia thesis. See p. 4.</td>
</tr>
<tr>
<td>CB</td>
<td>The common bases thesis. See p. 4.</td>
</tr>
<tr>
<td>DC</td>
<td>The disciplinary characterisation. See p. 23.</td>
</tr>
<tr>
<td>SA</td>
<td>The similarity approach. See p. 23.</td>
</tr>
<tr>
<td>HPA</td>
<td>The historical path approach. See p. 24.</td>
</tr>
<tr>
<td>DHP</td>
<td>Definition by the historical path. See p. 38.</td>
</tr>
<tr>
<td>CP</td>
<td>The core paradigm. See p. 54.</td>
</tr>
<tr>
<td>EH</td>
<td>The epistemic humility thesis. See p. 79.</td>
</tr>
<tr>
<td>CIF properties</td>
<td>Properties that are categorical properties, and/or intrinsic properties, and/or fundamental properties. See p. 80.</td>
</tr>
<tr>
<td>OSR</td>
<td>Ontic structural realism. See p. 107.</td>
</tr>
<tr>
<td>AMI</td>
<td>The argument from modal intuition. See p. 116.</td>
</tr>
<tr>
<td>HI</td>
<td>The Humean intuition. See p. 116.</td>
</tr>
<tr>
<td>CA</td>
<td>The categorical acquaintance thesis. See p. 166.</td>
</tr>
<tr>
<td>PPJ</td>
<td>The paradox of phenomenal judgment. See p. 175.</td>
</tr>
<tr>
<td>NCC</td>
<td>The neural correlates of consciousness. See p. 199.</td>
</tr>
<tr>
<td>The problem of possibility_{NCP}</td>
<td>The problem of the possibility of natural categorical properties. P. 231.</td>
</tr>
<tr>
<td>ENAMN</td>
<td>Epistemological naturalism (in accordance with metaphysical naturalism). See p. 259.</td>
</tr>
</tbody>
</table>
This page intentionally left blank
Introduction

Metaphysical naturalism (MN) is a doctrine, metaphysical framework, or worldview held or assumed by most contemporary philosophers and many of the folk. As David Papineau famously puts it, ‘[N]early everybody nowadays wants to be a “naturalist”’ (1993, p. 1). The doctrine presupposes that entities, properties, relations, and states of affairs can be categorised as ‘metaphysically natural’ or ‘metaphysically non-natural’ (also controversially called ‘supernatural’). It is roughly the following:

**MN:** The actual world contains only metaphysically natural things.

There is, however, no consensus among philosophers (and the folk) on what the metaphysically natural really is. Its proposed definitions in the literature are diverse, and every one of them is controversial. As John Dewey once wrote a century ago, ‘There is no word in the history of thought which carries more varied meanings than “nature”; naturalism shares in its diverse significations’ (2008/1927, p. 73). With this in mind, I will develop my preferred precise definitions of MN and the metaphysically natural in Part I. To help provide the reader with a very brief, preliminary understanding of MN (and the metaphysically natural) in the absence of a fully worked-out account, I will provide two examples of seminal definitions of MN offered by influential metaphysical naturalists. One is David Armstrong’s definition, according to which ‘reality consists of nothing but a single all-embracing spatio-temporal system’ (1978, p. 261); the other is John McDowell’s definition, according to which everything is enabled by laws of nature (1994, pp. 71-76). It should be obvious that the two definitions of MN above not only have tremendously different meanings, but also seem to fall under different philosophical topics. The situation is the same for other definitions of MN (or the metaphysically natural) in the literature. Despite the fact that MN (or the metaphysically natural) can be defined in such different ways, these philosophers believe they are engaging with the same subject, which is partially motivated by their agreement regarding which cases should be counted as metaphysically natural or non-natural.

According to conventional views, cases such as metal, the brain, ‘being organic’, and reproduction are metaphysically natural, while cases such as God, spirits, magic, and the fundamentally mental are not. Hence, major rivals of MN in contemporary philosophy are commonly considered to include substance dualism, idealism, panpsychism, theism, pantheism, and so forth. As
Kelly Clark points out, the rival views can be understood as the ‘spooky’ understandings of the world (2016, p. 2). More detail will be provided as this dissertation goes on.

Some readers might see similarities between MN, materialism, and physicalism. Indeed, the standard account is that materialism and physicalism are versions of MN. (For examples in philosophy of mind, see Armstrong 1978, 1995; Kim 2003; Guttenplan 1994; Melnyk 2009; Miłkowski 2008; Papineau 2015. For examples in philosophy of religion, see Moser & Yandell 2000; Taliaferro 2000; Kanzian 2006. For examples in metaethics, see Copp 2003; Crisp 2012. For other examples, see Dupré 2004; Dowell 2004; De Caro & Macarthur 2004.) In addition, MN is often considered a more modest doctrine than materialism or physicalism because everything material and physical must be metaphysically natural, while being metaphysically natural does not entail being material or physical. The most commonly recognised example of this fact is MN’s compatibility with emergentism in special sciences, which is the view that some irreducible ‘novel’ properties posited by special sciences ‘emerge’ out of more fundamental properties. While MN is often considered to be compatible with emergentism, materialism and physicalism are not (see for example, Armstrong 1978, pp. 269-270; Guttenplan 1994, p. 449; Kim 2003, pp. 88-89; Kanzian 2006, p. 91; De Caro & Macarthur 2007, p. 5; Melnyk 2009).

Some readers might also be reminded of methodological naturalism, a methodological doctrine according to which philosophy should be seen as an extension of science and should use similar methods. Even though some consider methodological and metaphysical naturalism to be two sides of the same coin (see, for example, Harman 2012, p. 10), this dissertation will focus on metaphysical naturalism and consider it as an independent doctrine. After all, there is a standard division between metaphysical naturalism and methodological naturalism (see, for example, Devitt 1998, p. 46; Papineau 2007, 2014; Horst 2009, pp. 221-222; De Caro & Macarthur 2010, p. 4; Clark 2015, pp. 1-5). As Papineau writes, ‘While both ontological and methodological naturalism claim a species of affinity between philosophy and science, the two doctrines are largely independent’ (2014, p. 116). No doubt, not everyone will agree with this division and my decision to focus on metaphysical naturalism. Those who are interested can refer to Appendix A, in which I will provide a detailed survey of the doctrines that the term ‘naturalism’ might stand for and a discussion of why they should be understood as separate.

As I have mentioned earlier, most contemporary philosophers take or assume MN to be true. This is unsurprising because there are a lot of attractive standard arguments in the literature that can support it against its major rivals (such as idealism, dualism, phenomenalism, theism, and so on).
They include the argument from physical causal closure in philosophy of mind, the problem of evil in philosophy of religion, the no-miracle argument in philosophy of science, and so forth. If these arguments are, as many philosophers believe, successful, then MN is very likely to be true and hence we are justified in believing it. This dissertation will take this likelihood as a prior probability and a default position.

The purpose of this dissertation is to investigate the connections between MN and the categorical ignorance thesis (CI), a metaphysical and epistemological thesis which has recently attracted attention through the work of Rae Langton (1998), David Lewis (2009), Frank Jackson (1998), and Simon Blackburn (1990), and to determine whether the latter will defeat our default position that the former is likely true and hence we are justified in believing it. CI can be defined as the following:

**CI:** Categorical properties are in principle unknowable. All of our knowledge is, in fact, about other properties, such as causal and dispositional properties.

What are categorical properties? They are properties that are not causal properties or dispositional properties (i.e. causal powers), and there are two kinds. The first kind are the structural properties that realise the dispositions we find in most cases in science. For example, the information output of a computer is realised by its complicated mechanical structure, and the dissolution of salt is realised by its molecular structure. The second kind is the kind of property that is non-structural, perfectly intrinsic, and fundamental, examples of which cannot be provided because current science (presumably) tells us nothing about such things. The categorical properties that CI and this dissertation are interested in include only the former. More detail will be provided in Part II.

The main purpose of this dissertation is to determine whether the ignorance of categorical properties, as Rae Langton (1998), John Foster (1993), Alyssa Ney (2007), Sam Cowling (2010), and David Braddon-Mitchell and Frank Jackson (2007) argue or worry about, is consistent with MN (or physicalism). To be more precise, this dissertation is concerned with the following two questions. (1) Is it possible for there to be a property that is unknowable, categorical, and, at the same time, metaphysically natural? (2) If yes, does CI imply a scepticism about the metaphysical naturalness of categorical properties, which will, in turn, significantly decrease the probability that MN is true? I believe that these worries are worthy of investigation because most if not all of the standard arguments
that motivate our default position (that MN is likely true and hence we are justified in believing it) are not intended to be used against them.

Russellian Monism (RM), an influential doctrine in philosophy of mind, will also be considered. The doctrine is as follows:

**RM:** (1) The scientific categorical ignorance thesis (SCI), (2) the categorical qualia thesis (CQ), and (3) the common bases thesis (CB) are true.

- **SCI:** Categorical properties are not knowable by natural science.
- **CQ:** Our qualitative experiences (i.e. qualia) are (solely or partially) constituted by their categorical bases (in some manner that is not solely via the causal, dispositional, and structural properties that are, in turn, borne by these bases).
- **CB:** Properties studied by science and our qualitative experiences share the same kind of categorical basis.

The doctrine is relevant to the topic because some proponents of it, such as Bertrand Russell (1992/1927a, 1927b), consider categorical properties to be knowable through our acquaintance with our qualitative experiences. On the other side, some philosophers, such as Daniel Stoljar (2001a, 2001b, 2006), David Chalmers (1996), and Galen Strawson (2013), argue that RM is compatible with MN (or physicalism). (Stoljar’s view merely implies that RM is compatible with MN because he only argues that RM is compatible with physicalism, which, as we have seen, is commonly considered a version of metaphysical naturalism.) This dissertation investigates whether the acquaintance view put forward by some Russellian Monists, understood as an objection to CI, is true and whether it can be compatible with metaphysical naturalism.

Some additional notes before we begin. Some readers may question why the main interest of this dissertation is MN instead of physicalism or materialism, which are similar to MN but have attracted more discussion. This choice is based on my belief that MN suits the purposes of this dissertation better than physicalism and materialism. This is for two reasons.

Firstly, committing to physicalism or materialism entails committing to MN, while committing to MN does not entail committing to either physicalism or materialism. This implies that, among the three doctrines, MN is held by the largest number of philosophers. Secondly, for the same reason, MN is considered the most inclusive and thus the most modest of the three. As everyone working in the discipline of philosophy knows, any view incompatible with physicalism, materialism, MN, or
similar views is strongly disfavoured by many philosophers. Some philosophers who are less sympathetic to mainstream metaphysical commitments than most make the move to deny physicalism and materialism while affirming MN (see, for example, Crane & Mellor 1995, p. 88; Chalmers 1996, p. 170). For example, Chalmers, one of the most influential writers on RM, writes, ‘To deny materialism is not to deny naturalism. A naturalistic dualism expands our view of the world, but it does not invoke the forces of darkness’ (1996, p. 170). (Note that he has equated physicalism with materialism.) Our major purpose here is to investigate whether the metaphysical framework shared by most contemporary analytic philosophers (which may be MN, physicalism, or materialism) can withstand the conceptual and epistemological challenge provided by CI; and one of the minor purposes is to investigate whether (the epistemology of knowing categorical properties offered by) RM is compatible with the framework. For these purposes, it is better to begin with the most widely held, inclusive, and modest framework.

One last point. There is a debate in the literature on whether MN should take abstract entities or properties (such as transcendent universals, platonic mathematical entities, timeless propositions, and so on) into account. To be more precise, the question is: can abstract entities or properties be categorised as metaphysically natural/non-natural, and is this relevant to whether MN is true? Some philosophers say yes to both (see, for example, Armstrong 1978; Moreland 2000), and others say the opposite (see, for example, Oppy 2014a). For the purpose of this dissertation, the abstract will be set aside, and the focus will be on the concrete.

The structure of this dissertation will be as follows. In Part I, I will further inquire into the nature of MN, with the purpose of discovering what it actually is. In Part II, I will determine whether CI is true and what the best argument(s) for it is. In Part III, I will provide a study of RM, which includes whether it is true, whether it allows for acquaintance of categorical properties, and whether it is compatible with MN. In Part IV, I will determine whether CI will defeat MN.
This page intentionally left blank
Part I

Metaphysical naturalism
This page intentionally left blank
Part I

Section 1. A method for characterising metaphysical naturalism

1.1. Introduction

The purpose of this dissertation is to inquire into the connection between metaphysical naturalism (MN) and the categorical ignorance thesis and determine whether the latter will challenge the former. To inquire into MN, we must first characterise or define it (and the metaphysically natural). There has been no lack of attempts to define MN in the literature, but rather than heading towards a resolution of the controversy, the debate seems to be heading in the opposite direction. The proposed definitions are tremendously different and none of them have been accepted by the majority. What this shows us is that philosophers understand the concept of MN in tremendously different ways and, in addition, lack a common ground to resolve their differences. With this in mind, it seems that three interrelated questions have to be answered. They are:

**The objectivity question:** Can a characterisation of MN be counted as correct by any objective standard?

**The methodological question:** Is there any method for characterising MN?

**The criteria question:** What are the criteria for a good characterisation of MN?

In this section, after a quick survey of some controversies surrounding the characterisation of MN, I will propose my method for characterising MN, which, if successful, will help us to answer all three questions at once. The method is, very roughly, to provide a regimentation of a shared belief held by the philosophical community. Details of the method will be provided as the discussion goes on. The definition(s) of MN in Section 2 will be developed by the method. Consequently, it is noteworthy that this section is not about what MN is; rather, it is about the method for discovering what it is.

The structure of this section will be as follows. In Section 1.2, I will provide an overview of the efforts of philosophers to define MN in the literature. In Section 1.3, I will outline the difficulties in defining MN and suggest how we can overcome them. In Section 1.4, I will propose a method for defining MN. In Section 1.5, I will provide some criteria for a good definition of MN under the
guidance of the method. Since this method, as we will see, takes into account current consensus on whether something is considered metaphysically natural, in Section 1.6 I will provide a special consideration of the fundamentally mental because there is controversy surrounding them with regard to this question.

1.2. Previous efforts by philosophers to define MN

Before beginning my discussion of the method for characterising MN, I will first provide a brief overview of the efforts of philosophers to define MN and the definitions that have been offered, in order to provide the background of the debate. Despite the fact that MN is the dominant doctrine held or assumed by most contemporary philosophers, there is a lack of consensus as to what the metaphysically natural is. The controversy concerning which definition of MN is correct is pretty intense. The reason, I believe, is that while MN is considered by most to have a strong association with natural science, contemporary scientific theories are growing increasingly complex and abstract. In particular, it is not easy to see what is in common between the entities or properties posited by scientific theories, which the definition of the metaphysically natural could make use of. For example, in contemporary physics, physical entities are no longer understood as necessarily spatiotemporal, and hence any conception of the metaphysically natural that relies on spatiotemporality (see, for example, Armstrong 1978) has unavoidably become unattractive. While philosophers seem to share similar conceptions of what the metaphysically natural is – especially when it comes to whether a particular case, such as God or a chair, is metaphysically natural – controversies always arise when they try to spell out these conceptions.

In the literature, there is no lack of attempts to develop a precise definition of MN or the metaphysically natural. Some define the metaphysically natural by specifying some abstract features of metaphysically natural entities, e.g. being spatiotemporal or dependent on the laws of nature (see, for example, Armstrong 1978; McDowell 1994; Chalmers 1996; Devitt 1998; Schellenberg 2007). Some define the metaphysically natural via negativa, e.g. not being a fundamental mentality, or a telos, or worthy of worship (see, for example, van Inwagen 2006; Oppy 2014a). Some define the metaphysically natural with reference to theories in current science (see, for example, Chalmers 1996). Some define the metaphysically natural with reference to theories in ideal science (see, for example, Miłkowski 2008). Some define the metaphysically natural using an ostensive definition, i.e. define the metaphysically natural by pointing out examples like chairs and tables (see, for example, Searle 2007). Despite the fact that all these efforts have been made, there is no sign that the debate is heading
towards a resolution of the controversy. None of these definitions have gained wide acceptance. On the contrary, what the debate has shown us is only how tremendously differently philosophers understand the concept of the metaphysically natural and how unlikely it is that these differences will be resolved.

1.3. Difficulties in defining MN

We have seen the diversity of definitions of MN in the literature. A worry that will unavoidably arise is that the term ‘metaphysical naturalism’ may simply have many meanings, and that there is no single correct definition of it. This relates back to the objectivity question I mentioned earlier. Recall that it is the following:

The objectivity question: Can a characterisation of MN be counted as correct by any objective standard?

In this subsection, I attempt to provide an answer to the question by saying ‘Yes!’ and explaining why.

We shall begin by going through the worry in detail before discussing my answer. According to Mario De Caro and Alberto Voltolini’s interpretation of David Papineau’s article ‘Naturalism’ (2015) in the Stanford Encyclopedia of Philosophy, Papineau argues that ‘it would be fruitless to try to adjudicate some official way of understanding the term’ because ‘different contemporary philosophers interpret the term differently’ (De Caro & Voltolini 2010, p. 71). In addition, as Papineau himself notes, few philosophers are happy to be considered as ‘non-naturalists’ nowadays, and hence different philosophers often interpret the term differently so that they will not be disqualified as naturalists. He writes:

This inevitably leads to a divergence in understanding the requirements of ‘naturalism’. Those philosophers with relatively weak naturalist commitments are inclined to understand ‘naturalism’ in an unrestricted way, in order not to disqualify themselves as ‘naturalists’,

---

1 It is noteworthy that it is unclear that Papineau is really suggesting that there is no single correct definition of the term, as De Caro and Voltolini interpret him as suggesting. Papineau’s article ‘Naturalism’ (2015) is an encyclopedia article in the Stanford Encyclopedia of Philosophy, thus he might merely be suggesting that there is not a single neutral, non-opinionated and non-controversial way to introduce the use of the term in the contemporary philosophical discipline, not that there cannot be one single correct definition if we understand the term in an opinionated way. I take no stance on this matter.
while those who uphold stronger naturalist doctrines are happy to set the bar for ‘naturalism’ higher. (2015)

What (De Caro and Voltolini’s) Papineau seems to provide is a negative response to the objectivity question. More precisely, his argument seems to be that since there is no agreed standard of whether a characterisation of MN can be counted as correct in the literature, there is probably no objective standard either.

Even though I do not agree with him, I find his argument appealing and worrisome. Indeed, given the diversity of interpretations of the term ‘metaphysical naturalism’, its definition might simply be a linguistic matter, which implies that ‘metaphysical naturalism’ is merely a term with multiple linguistic meanings. As with many ordinary words in the dictionary, it is not obvious that there is only one single correct definition of MN.

This leads to a significant challenge to the purpose of this dissertation. The purpose of this dissertation is to assess the justifiability of the widely held metaphysical framework in contemporary philosophy, MN, in the face of the challenge from the categorical ignorance thesis. If the concept of MN is imprecise or even indefinable, and every philosopher means a different thing when she uses the term, then we may have to accept the fact that there is no such widely held metaphysical framework in the first place, since what the philosophers commonly hold might merely be a term rather than a metaphysical framework. On the other hand, even if there can be a so-called objective standard of whether a characterisation of MN can be counted as correct or not, it may be irrelevant to most philosophers who are typically categorised as metaphysical naturalists. This is equally damaging to our purpose.

But the situation is not as hopeless as it seems. The fact that there is no agreed standard does not imply that there is not an objective one which is relevant to most metaphysical naturalists, but it is evidence that no such standard exists. Like all evidence, its force can be countered by evidence for the opposite position. And in our case, the best way to discover such evidence for the opposite position is to take up the task to develop an objective standard as required for our purposes. Therefore, what I am going to do is to focus on answering the methodological question, i.e. to develop a method for characterising MN. Such a method, as far as I am concerned, is currently lacking in the literature on defining MN (and even in the literature on defining its relative, physicalism). If my method turns out to be promising, then we at least have a hope of determining an ideal definition of MN in some
objective manner, and the objectivity question and (De Caro and Voltolini’s) Papineau’s challenge will naturally vanish.

1.4. A method for characterising MN

1.4.1. The method

We have seen that (De Caro and Voltolini’s) Papineau’s challenge is appealing, and it is natural to fall into the grip of the pessimistic belief that there is no possible solution to it. However, among philosophers’ work on characterising folk concepts like <qualia>, <free will>, <political freedom>, <democracy>, and so on, these difficulties are, in fact, not uncommonly seen. Benefitting from the long-term experience of philosophers in dealing with concepts like these, there are already some good approaches to characterising these concepts as precise philosophical doctrines in the literature. One example provided by David Lewis (1995) is his approach to naturalising qualia. Lewis teaches us that we can see a folk concept as a functional role spelled out in our tacitly held folk theory, and then search for the best deserver of that functional role, even though there might not be a perfect deserver of it.

Certainly, the Lewisian method cannot be applied directly to the case of characterising MN due to three important differences. Firstly, in Lewis’s case, deservers are objects, properties, or states of affairs, and, in addition, they have causal powers that are parallel to the functions possessed by their roles. In the case of characterising MN, what is required is not a deserver in the Lewisian sense but (a definition of) a concept.

Secondly, the concept of metaphysical naturalness is not merely a folk one, since its meaning has been altered and developed in its use by philosophers, who are sophisticated and self-reflective theorists. Hence, it is inappropriate to assume that it remains entirely intuitive. While the folk talk about things being natural and supernatural, philosophers have borrowed these concepts and developed them into more sophisticated versions in their philosophical work. For example, while werewolves, aliens, and UFOs are often considered by the folk to be supernatural, I doubt many philosophers share these views. In fact, among these philosophers, some are even very clear about what they themselves mean by ‘metaphysical naturalism’. This is seen in the cases I surveyed in Section 1.2.

The third difference from Lewis’s case follows from this, and it is perhaps the most important difference. The vagueness, imprecision, and perhaps inconsistency of the term ‘metaphysical
naturalism’ do not arise at the individual level but at the group level. As I mentioned, the group of people who use the term ‘metaphysical naturalism’ are sophisticated and self-reflective theorists, and are thus not unreflective and guided by primitive intuitions like the folk. It is not the case that what they themselves take to be MN is unclear. However, when these theorists attempt to refer to MN, they often refer to a shared concept the philosophical community has in mind rather than one of their own. For example, in the standard context of philosophy of religion, the scientific atheist camp is commonly referred to as ‘the naturalists’ by both theists and atheists. This shared understanding cannot be a definition an individual philosopher has in mind. The obscurity, imprecision, irregularity, and perhaps inconsistency arise here, since it is never easy to make clear what the content of a shared concept is, even if the shared concept is exhaustively shared by experts. This is because none of them can speak for the whole community, no matter how much of an expert they are. This is why each attempt by an individual philosopher to define MN has attracted disagreements from so many others, metaphysical naturalists and non-naturalists included.

Nonetheless, an analogue of Lewis’s method is available in this case. Rather than searching for the best deserver of a folk-theoretical role, we can search for the best regimentation of a shared belief. In the case of MN, it is an unspecified doctrine that consists of a collection of theoretical roles (such as being incompatible with theism, having a close connection to contemporary science, etc.) that are spelled out in some consensus tacitly held by a large group of experts. While a doctrine in this form is presumably unsystematic and inconsistent, its regimentation is supposed to fulfil most of the roles which the original consisted of, while being clearly specified, systematic, and consistent. As with Lewis’s method, there might not be a perfect regimentation because it is possible that no regimentation can fulfil all of the roles while remaining systematic, consistent, and reasonable, but we can still look for a best one.

I consider the characterisation of MN formulated by this method to be not only useful in general terms, but also useful to the purpose of this dissertation. This is because this formulated characterisation will not only be my own version of MN, but also a good representation of the shared belief, contemporary MN, which many contemporary philosophers identify themselves with. Being able to spell out this belief for philosophical assessment is useful in its own right because it allows us to reflect on the shared theoretical commitment. The purpose of this dissertation is such an effort, namely to assess whether the shared belief can withstand the challenge provided by the categorical ignorance thesis.
1.4.2. Does the tacit consensus exist?

We now have a Lewisian-like method for characterising MN. The problem is then whether a tacit consensus concerning what MN is exists. If the tacit consensus does not exist, then the method cannot be applied. Fortunately, I think the answer is ‘yes’. My argument is as follows.

Although there is widespread disagreement among metaphysical naturalists on what MN actually is, frequent uses of the term ‘metaphysical naturalism’ by philosophers (and the folk) do not often create a lot of confusion among their users, including its proponents and opponents, and the relevant debates can often go on without distortion. This would not be possible if there was nothing common and consensual in their uses of these terms. There must be some tacit consensus in the literature on what MN should be like and/or which particular things are metaphysically natural.

It is noteworthy that this tacit consensus can come about in two ways. The first is a theoretical conception of what MN itself is; the second is an association of the term ‘metaphysical naturalism’ with the view(s) of some particular individual(s), in the same way that, say, Kantianism is associated with Immanuel Kant or existentialism is (to a certain extent) associated with Jean-Paul Sartre. What allows philosophers to talk about MN without causing confusion in the second case is their shared tendency to think of the same writer(s) when they use the term. I will argue that, while I speculate that MN is an example of the first case, our method applies to both cases.

I speculate that MN is an example of the first case but not the second case because it does not seem to me that there is any philosopher that can be said to exhaustively represent MN. Certainly, there are influential contemporary figures in this field, such as J. J. C. Smart, David Lewis, David Armstrong, Frank Jackson, Jaegwon Kim, Jerry Fodor, Paul and Patricia Churchland, John McDowell, and so on, as well as influential historical figures, such as Thomas Hobbes, Julien La Mettrie, Baron d’Holbach, and so forth. But being influential in the debate on a doctrine or theory is one thing, and representing the doctrine or theory is another. It is not obvious that, as far as I am concerned, any of these figures, or any combination of them, can be seen as exhaustively representing MN or physicalism in the way Kant represents Kantianism or the way Sartre represents existentialism.

Regardless, neither kind of tacit consensus is in conflict with our method. This is because the difference is only whether the term ‘metaphysical naturalism’ in standard philosophical discourses refers to a specific set of ideas or the ideas of a specific person, and in either case we can determine what the ideas are. As long as we can determine what those ideas are, then we can (try to) provide a regimentation.
1.4.3. Answering the three crucial questions

With our method for characterising MN, we can answer the three questions raised at the beginning. Recall that they are:

**The objectivity question:** Can a characterisation of MN be counted as correct by any objective standard?

**The methodological question:** Is there any method for characterising MN?

**The criteria question:** What are the criteria for a good characterisation of MN?

The methodological question is directly answered by the existence of our method. And according to our method, a good characterisation of MN is a good regimentation of the tacit consensus with regard to what the term ‘metaphysical naturalism’ is taken to mean. Hence, the criteria question can be answered as follows: the criteria of a good characterisation of MN are the criteria of a good regimentation, which I will develop in Section 1.5.

The objectivity question is also implicitly answered. What the consensus really is and what its best regimentation is are both objective matters. No doubt, the boundary between the objective and the subjective is unclear. A tacit consensus is perhaps better understood as an intersubjective matter rather than an objective one. Nevertheless, setting the metaphysical nature of consensus aside, it is obvious that there is an objective answer with regard to what a particular consensus is. After all, objective facts about semantics are always facts about the linguistic habits of a community.

1.5. Criteria for a good regimentation of MN

The method I proposed in Section 1.4 is an abstract guide for searching for a characterisation of MN. What we are now in need of are some precise criteria for a good regimentation of MN, since we do not want the comparison between possible regimentations to be arbitrary. Here I will suggest several criteria I consider useful.

I believe that the criteria can be classified into two general types, which are (1) general theoretical virtues and (2) virtues specific to regimenting MN.

1.5.1 General theoretical virtues

*General theoretical virtues* are those theoretical virtues that philosophical work in general requires, and regimentations of MN are no exception. Examples include logical consistency, linguistic clarity,
and some basic epistemic virtues, such as simplicity, non-circularity, non-arbitrariness, non-trivialness, fruitfulness in suggesting future work, and so on. Since these virtues are generally agreed upon by philosophers and are considered by most to be the basic requirements of any philosophical work, I will suppose that I can reasonably assume them to be correct and sufficiently known by the reader. Hence, I shall not discuss them any further.

1.5.2 Virtues specific to regimenting MN

Contrary to general theoretical virtues, virtues specific to regimenting MN are not virtues that philosophical work generally requires but those that are specifically required in regimenting MN. As I have argued, a good regimentation of MN should be able to fulfil most if not all roles spelled out by the tacit consensus concerning what MN is like. These roles, I believe, include the agreed cases and the tacit theories. The criteria of a good regimentation therefore include (a) consistency with tacit theories and (b) consistency with agreed cases. I also add an additional criterion, (c) having fewer uncommon or idiosyncratic presuppositions, among virtues specific to regimenting MN for the purpose of this dissertation. I will explain the reason below as I discuss the virtue.

1.5.2.1 Consistency with tacit theories

There seem to be some tacit theories shared among philosophers in philosophical debates about what MN should and should not be like. It is not easy to state all tacit theories explicitly because tacit theories are by nature implicit, but some obvious ones are not difficult to discern and can be taken as examples. These include that MN is a metaphysical and ontological thesis, that it at least shares some similarities with physicalism, that it has a close relationship to discoveries in modern natural science, etc.

Furthermore, many philosophers’ definitions of MN seem to be attempts to develop, systematise, and formalise these tacit theories into precise and useful philosophical concepts. Even though I might not agree that these definitions are the best deservers of the tacit theories, I think that they are valuable opinions on what the tacit theories are like. Therefore, a good regimentation of MN should not have too many inconsistencies with these proposed definitions.

1.5.2.2. Consistency with agreed cases

Two of the agreed theoretical roles of MN are the exclusion of some agreed cases of the metaphysically non-natural and the inclusion of some agreed cases of the metaphysically natural.
Although there is no consensus on what MN is, there is a strong consensus on the naturalness and non-naturalness of certain things.

Agreed cases of the non-natural: As Jaegwon Kim argues, following Roy Wood Sellars, minds and mental properties are a crucial test case for metaphysical naturalists’ attempts to define naturalism (2003, p. 96). Fundamental mentalities that are not reducible to or realised by (or emergent from) lower-level properties are generally considered to be metaphysically non-natural. Other cases include God, gods, curses, karma, etc.²

There are some controversial or less obvious cases, such as numbers and universals. As they are not agreed cases, I think that one should determine whether they are natural after choosing a definition that one prefers. However, as I have mentioned in the Introduction, I shall not provide any further discussion of abstract objects, as they are outside the scope of this dissertation.

Agreed cases of the natural: Most physical objects, chemicals, and biological structures are commonly considered to be metaphysically natural. Moreover, while string theory in physics and some discoveries and theories in quantum physics are highly counterintuitive, most metaphysical naturalists consider them naturalistically respectable.³ A good regimentation of MN should be able to count all of these things as metaphysically natural, however difficult that may be.

1.5.2.3. Having fewer uncommon or idiosyncratic presuppositions

As I said earlier, the principle of having fewer uncommon or idiosyncratic presuppositions is an additional criterion that is specific to the purpose of this dissertation. I will explain the reason for this criterion as the discussion goes on. Let us begin with an introduction to the principle. When formulating a definition of MN, some philosophical presuppositions might be made within the definition. For example, a definition of MN might make use of a particular interpretation of the history of science, such as the Kuhnian one or the Popperian one, or a particular view on laws of nature, such as the Lewisian one or the Russelian one, or a particular view on causation, such as the Humean one or the counterfactualist one. However, MN is supposed to be a fairly abstract philosophical position that does not assume its believers to have many uncommon or idiosyncratic background philosophical

² Here I set aside whether some cases overlap with each other. For example, many consider God and gods to have minds, even though this is controversial because many theists do not believe in a personal God (i.e. a God with a personality like ours).
³ When I say a theory is ‘naturalistically respectable’, I mean its contents are all compatible with metaphysical naturalism. This is different from saying that the theory is metaphysically natural, which might be understood by some to mean that the theory itself, as an abstract entity, is metaphysically natural.
presuppositions. The principle of having fewer uncommon or idiosyncratic presuppositions thus suggests that such presuppositions should be avoided as much as possible.

Certainly, believing in a doctrine does not require one to believe in the philosophical presuppositions within its definition. For example, even if a definition of MN presupposes that the Popperian interpretation of the history of science is true, a Kuhnian can still make use of the definition and find out that her views are naturalistically respectable. This is because the Kuhnian can assume that the Popperian interpretation of the history of science is true when she uses the definition. The assumption does not have to be considered true, since it can just be assumed for methodological or pragmatic purposes. This way of using assumptions is not implausible, and is, in addition, commonly seen in studies in natural science. In particular, many mathematical formulas in scientific theories are assumed in the same way. For example, most, if not all, ecologists know that ecosystems are in some way chaotic because they are composed of individual organisms whose behaviours and fates are arbitrary, but abstract mathematical formulas are still assumed because of their usefulness in predicting results. However, the worry here is that the Kuhnian mentioned, for example, might not find the definition meaningful because of its groundlessness, whether or not she can use it. This is because it is based on a philosophical presupposition that she cannot accept. In opposition to this, MN is not supposed to be a view meaningful only if one is committed to very specific philosophical views.

Some might believe that the principle of consistency with tacit theories already presupposes the principle of having fewer uncommon or idiosyncratic presuppositions because the latter is based on the tacit assumption that MN is fairly abstract. I agree, but there is a pragmatic reason for specifically and separately listing it. Recall that the purpose of this dissertation is to assess the justifiability of MN, the doctrine or the metaphysical framework commonly held or assumed by many contemporary philosophers, in the face of challenges provided by the categorical ignorance thesis. In line with this purpose, I hope that the kind of MN being assessed is one that most metaphysical naturalists could find relatable, and this requires that the definition we assess to be one that most find meaningful. If the kind of MN under consideration was not meaningful to most, many would simply consider it to be a doctrine that was irrelevant to their work and thoughts. Therefore, I believe that there should be an extra consideration with regard to the principle of having fewer uncommon or idiosyncratic presuppositions.
A possible objection to my set of criteria is worthy of discussion before we go on. J. L. Dowell, in her article on physicalism, argues that whether a particular thing is physical or not should be an *a posteriori* and fallible issue (2006, pp. 42-45). The same can presumably be said about metaphysical naturalism. For example, it should be possible for us to discover that a tree is not really metaphysically natural because of new empirical discoveries about its nature. Some might consider the principle of consistency with agreed cases to be in conflict with Dowell’s principle. The reason is as follows.

If the formulation of a definition is case-based, the current consensus on the categorisation of metaphysically natural and non-natural things would be an *a priori* and infallible truth. For example, the fact that a cup is metaphysically natural would be *a priori* and infallible. This is because any definition formulated in a case-based way will have to be formulated according to the presupposition that a cup is metaphysically natural, and hence it will be in principle impossible for us to discover that the case is otherwise. In sum, according to the objector, the principle of consistency with agreed cases does not leave space for the possibility that the current consensus might be wrong. This will be a shortcoming if we agree with the objector that whether a thing is metaphysically natural should be an *a posteriori* and fallible issue.

What is the alternative method according to the objector? One possibility is the following. Rather than having a case-based method, the formulation of a definition can be merely theory-based, i.e. formulated in terms of theoretical considerations without considering the currently agreed cases. After the correct theory of MN is formulated, we can then categorise things as being metaphysically natural or non-natural through empirical investigations of their nature. This allows our current categorisation to turn out to be wrong.

I agree with Dowell and the objector that the categorisation of whether a thing is metaphysically natural should be an *a posteriori* and fallible issue. However, it is important to note that my method is neither merely case-based nor merely theory-based. This is because I have included both the principle of consistency with tacit theories and the principle of consistency with agreed cases among the criteria for a good regimentation of MN. In addition, there is nothing *a priori* and infallible in these principles because they are not considered as inviolable principles that are to be respected and followed at all costs. With this in mind, my method is similar to John Rawls’s (1971) influential method in moral and political philosophy, the method of reflective equilibrium. It goes back and forth between the theory and the judgments of cases, and revises both of them until an equilibrium is
reached, i.e. the theory and the judgments of cases finally coincide (p. 20). With this method, the categorisation of whether a thing is metaphysically natural is not an *a priori* and infallible truth since it can be revised in the procedure of reflective equilibrium.

No doubt, the objector may still question why we cannot have a method that has the principle of consistency with tacit theories alone, since this will make our task simpler. My response is that this is impractical. As we have seen, ‘metaphysical naturalism’ is a name for an imprecise, vague, and often intuitive concept. All that can be discovered is a best regimentation of the concept. The tacit theories are, still, very imprecise and hence cannot possibly suggest a precise definition alone. Kim’s suggestion to have test cases is far more realistic. In addition, as Frank Jackson (1998, pp. 31-32) suggests, the attitudes used in the judgements of cases can actually reveal further tacit theories that cannot be discovered by theory-based methods. To a certain extent, the principle of consistency with agreed cases can be understood as being implied by the principle of consistency with tacit theories (supposing tacit theories are understood in a broad sense), which includes the hidden attitudes involved in the judgements of cases.

**1.6. The controversy surrounding the fundamentally mental**

I have listed my criteria for a good regimentation of MN, which include the principle of consistency with agreed cases. Although there is a general consensus about which things are metaphysically natural and which things are not, this does not imply that disagreements do not exist. Before starting to formulate my definition using the criteria outlined above, controversial cases should be considered to make sure that the principle of consistency with agreed cases can be applied properly. With this in mind, the controversy surrounding the fundamentally mental needs to be considered.

While most hold the assumption that the fundamentally mental must be metaphysically non-natural, a few philosophers have disagreed with this common assumption. Two influential writers on panpsychism, David Chalmers (1996) and Galen Strawson (2013), both claim to be metaphysical naturalists and both suggest that some versions of panpsychism are naturalistically respectable.

My view is that it is quite obvious that, according to the consensus, the fundamentally mental cannot be metaphysically natural. Philosophers like Chalmers and Strawson, it seems to me, are still a minority with radical views, and they seem to be attempting to disagree with the general consensus about how the term ‘metaphysical naturalism’ should be understood and to be offering a revision of it. Since their views are not part the general consensus, I will not consider them as I develop my characterisation of MN in Section 2.
It is important to note that being in the minority, certainly, does not imply that their suggestions for revision are incorrect. My method does not rule out such a possibility. The exclusion of the fundamentally mental is merely an input to the formulation of the definition of MN; whether it will turn out be tenable is another matter, since the resulting definition will be derived through reflective equilibrium. Through reflective equilibrium, it might, after a series of case-based and theory-based considerations, be concluded that the exclusion of the fundamentally mental should indeed be given up. If this turns out to be the case, I will not attempt to add any ad hoc condition into the definition in order to make it exclusive of the fundamentally mental, because simplicity is a general theoretical virtue I consider important (see Section 1.5.1). For example, if David Armstrong’s definition of MN, according to which ‘reality consists of nothing but a single all-embracing spatio-temporal system’ (1978, p. 261), had turned out to be the best regimentation of MN we can discover and it cannot exclude the fundamentally mental, I would not adjust it to ‘reality exhaustively consists of a single all-embracing spatio-temporal system and no fundamentally mental things’ without any good reason.

In sum, whether the fundamentally mental should finally be considered as metaphysically natural rests on whether a reasonable and simple definition of MN that excludes them can be formulated. If they cannot be excluded without the addition of ad hoc conditions, then I will choose the definition without such exclusion and will come to be in agreement with Chalmers and Strawson.

1.7. Conclusion

I have provided a method to characterise MN, which is to see it as collection of theoretical roles spelled out in some tacit consensus within standard philosophical discourse, and then to provide a regimentation of it. This provides us a way to assess MN as a doctrine or metaphysical framework held by most contemporary philosophers despite the fact that there is no consensus of what it is. The three criteria for a good regimentation I developed through the method are (a) consistency with tacit theories and (b) consistency with agreed cases and (c) having fewer uncommon or idiosyncratic presuppositions, which I will make use of when I develop my preferred definition of MN in Section 2.
Part I

Section 2. Characterising metaphysical naturalism

2.1. Introduction

The purpose of this dissertation is to inquire into the connection between metaphysical naturalism (MN) and the categorical ignorance thesis and determine whether the latter will challenge the former, and hence we have to determine what MN actually is. In Section 1, I outlined a method for characterising MN. In this section, I will actually propose a definition.\(^4\) Given the criteria listed in Section 1, I consider the most popular approach to defining MN and physicalism, the *disciplinary characterisation* (DC), to be the (apparently) most attractive one.\(^5\) It defines MN with reference to scientific theories and defines physicalism with reference to physical theories, and suggests that every entity or property that exists is posited by those theories. But it has been widely criticised for its inability to solve (1) Hempel’s dilemma, (2) the problem of naturalistically respectable theories (or theorists) in history, (3) the problem of non-actualised properties and possible worlds containing them, (4) the problem of the scientifically inaccessible, and (5) the problem of naturalistically unacceptable scientific theories.

In this section, I will introduce DC and the main objections to it. I will then introduce a seminal approach to DC that I call the *similarity approach* (SA), which is defended by David Lewis (1983), Ian Ravenscroft (1997), and David Braddon-Mitchell and Frank Jackson (2007). The suggestion is to allow an extension of current science, constrained by an adequate similarity to it. I will argue that while it can avoid most of the main objections to DC, it has a fatal shortcoming that I call the *dilemma of the right kind of similarity*. The concern is that, for there to be a level of similarity, precise or not, that acts as a boundary between the metaphysically natural and non-natural, we need to have the right kind of similarity as the measure. Our available options include overall similarity and similarity in a certain respect, but I will argue for the following. On the one hand, given what we expect the boundary

\(^4\) Henceforth, I will use the terms ‘definition’ and ‘characterisation’ interchangeably, because their differences (if there are any) are irrelevant to my purpose.

\(^5\) For the origin of the term, see Copp 2012. David Copp himself does not endorse this characterisation and, instead, believes that we should abandon it (2012, p. 28). Nevertheless, I consider his term ‘disciplinary characterisation’ a good title of the approach.
between natural and non-natural to be, overall similarity obviously cannot work. On the other hand, accepting similarity in a certain respect is in fact an abandonment of the SA.

As such, I will develop an original approach to DC using the strategy of SA, according to which extensions of current science are allowed. I call it the *historical path approach* (HPA). The idea is (roughly) that MN can be defined with reference to the historical ideas that current scientific theories descend from. With this, I will argue, we can avoid the dilemma of the right kind of similarity. Finally, I will show that a definition of MN developed by HPA is a useful definition and can be developed into a substantive metaphysical doctrine despite its reliance on sociohistorical considerations.

I suggested in Section 1 that what we should aim for is the best regimentation of the consensus of the philosophical community with respect to what MN is like. I hope the reader will find my definition of MN the best one after reading this section.

### 2.2. Introduction to the disciplinary characterisation

As we have seen in Section 1.2, despite the common uses of the concept of MN in different fields of contemporary philosophy – including philosophy of mind, philosophy of science, metaphysics, metaethics, philosophy of religion, philosophy of language, etc. – the proposed definitions of it in the literature are diverse. Of them, the ‘disciplinary characterisation’ (DC) is the most popular one among philosophers. If we take not only MN, but also physicalism and materialism into account, then its proponents even include influential theorists such as Smart (1978), Lewis (1983), Jackson (Braddon-Mitchell & Jackson 2007), and many others. Roughly, the DCs of MN and physicalism are:

**DC (MN):** Every entity or property instantiated in the actual world is metaphysically natural – in the sense of being posited by natural science or being exclusively constituted by (i.e. reducible to or realised by) those posits.

**DC (physicalism):** Every entity or property instantiated in the actual world is physical – in the sense of being posited by physics or being exclusively constituted by (i.e. reducible to or realised by) those posits.

It is noteworthy that MN and physicalism as characterised by DC are not the same. It is not uncommon among philosophers to consider oneself to be a metaphysical naturalist but not a physicalist (see, for example, Crane & Mellor 1995, p. 88; Chalmers 1996, p. 170), and DC allows this kind of categorisation. Physicalism is characterised by physics only, while MN is characterised
not only by physics, but also by natural science in general, including chemistry, molecular and cellular biology, ecology, geology, and so on. Hence, to be a physicalist, one has to believe that entities or properties posited by those scientific disciplines (but not physics), such as bacteria, cats, forests, and so on, either (1) are exclusively constituted by (i.e. reducible to or realised by) entities or properties posited by physics or (2) can be eliminated. MN, on the contrary, does not have this constraint. This is because the higher-level entities or properties posited by special sciences are directly referred to, and many scientists believe that they are emergent entities or properties. With this in mind, they do not need to be exclusively constituted by (i.e. reducible to or realised by) entities or properties posited by physics, even if they are not posited by physics. They can be, as emergentism suggests, emergent entities with properties that are ‘novel’. Hence, many philosophers consider the compatibility with emergentism in special sciences as a notable example of what distinguishes MN from physicalism (see, for example, Guttenplan 1994, p. 449; Melnyk 2009; De Caro & Macarthur 2010, p. 5). This dissertation is about MN and not physicalism. However, I take it that most of the arguments in this section apply equally to both doctrines, so I will also consider arguments by authors concerning the definition of physicalism. Henceforth, I will focus on MN, but when I discuss arguments by an author who is discussing physicalism or arguments that target physicalism only, I will discuss physicalism.

It is also noteworthy that some readers might be reminded of methodological naturalism because of the title ‘naturalism’ and the way DC characterises it. Recall that methodological naturalism is a methodological doctrine, according to which philosophy should be seen as an extension of science and should therefore use similar methods. I have noted in the Introduction that methodological naturalism will not be considered in this dissertation, but in the case of DC, some readers might find my decision unconvincing because of the strong relation with science shared by DC and methodological naturalism. Nevertheless, in the literature, there is a standard division between MN and methodological naturalism, even in the cases where MN is defined by DC (see, for example, Papineau 2007; 2014, pp. 221-222; De Caro & Macarthur 2010, p. 4). As David Papineau puts it:

This ‘methodological naturalism’ is to be distinguished from ‘ontological naturalism’, understood as a general view about the contents of reality. Ontological naturalism maintains that reality involves nothing more than the entities studied in the natural sciences and contains no supernatural or transcendent realm. While both ontological and methodological naturalism claim a species of affinity between philosophy and science, the two doctrines are largely independent. (2014, p. 116)
In the rest of this section, I will assume this standard division and set methodological naturalism aside. Those who disagree (and those who are interested) can refer to Appendix A, in which I will discuss the variety of doctrines that the term ‘naturalism’ might stand for in more detail.

2.3. The prima facie strengths and weaknesses of DC

DC is both intuitively attractive to many, myself included, and has some prima facie explanatory power and simplicity. It appears to somehow match with the principle of consistency with tacit theories, the principle that a definition of MN should be consistent with the tacit theories shared among philosophers with regard to the nature of MN (see Section 1.5.2.1). This is because, as would be the case were DC true, we seem to distinguish whether something is metaphysically natural or not by its relation to natural science. In addition, DC seems to have no obvious inconsistency with our other tacit theories. One exception is that some might consider what scientists believe to be a sociohistorical matter, not an issue of metaphysics, and this, in turn, might be in conflict with the tacit theory that MN is a metaphysical and ontological doctrine. I will return to this problem later on.

It also appears that DC matches well with the principle of consistency with agreed cases, the principle that a definition of MN should be consistent with our judgments concerning the metaphysical naturalness of each particular entity of property (see Section 1.5.2.2). Consider cases that most take into consideration, such as tables, trees, Cartesian minds, God, and so on. It seems that they can all be very much distinguished into their agreed categories by DC. Furthermore, it has no apparent conflict with the principle of having fewer uncommon or idiosyncratic presuppositions, the principle that a definition of MN should avoid philosophical presuppositions that are uncommon or idiosyncratic (see Section 1.5.2.2).

This provides an advantage in comparison with some other proposed definitions in the literature because most of them either are radically inconsistent with the agreed cases or are cluster definitions that are arbitrary and ad hoc. Let us begin with the non-cluster ones. It is impossible to assess all of them here, but we can quickly go through two notable examples, David Armstrong and Michael Tye’s formulations of naturalism. Consider David Armstrong’s seminal formulation of naturalism:

---

6 Some friends of the creationist movement and the New Age movement might disagree, because they believe that their studies of God’s creation and spirituality are legitimate scientific studies. But these views are typically religiously motivated and are rejected by the mainstream academic community. Hence, in the rest of this section, I shall set them aside.
Reality consists of nothing but a single all-embracing spatio-temporal system. (1978, p. 261)

This definition has some obvious inconsistencies with the agreed cases. A spatiotemporal God has to be counted as metaphysically natural, while a physical theory that does not take spatiotemporality to be real or fundamental cannot be counted as naturalistically respectable. Consider Michael Tye’s formulation of naturalism:

Higher-level naturalistic phenomena (both token and type) are either ultimately constituted or ultimately realized by microphysical phenomena. (1995, p. 42)

Much like Armstrong’s formulation, Tye’s formulation seems to be inconsistent with some agreed cases. While microphysicalism might be an assumption of the old schools of physics, it is no longer the case. Both current physics and current philosophy of physics no longer see the commitment to the microphysical as their dogma. For example, it is not obvious that a quantum field is small-sized. Furthermore, according to quantum holism, which many are attracted to, particles are, in fact, not independent entities but a single holistic entity (see, for examples of relevant discussions, Teller 1986; Oppy 2000; Schaffer 2003; Esfeld 2004; Ladyman and Ross 2007; Papineau 2008). As far as I am concerned, no one, metaphysical non-naturalists included, has raised any doubt on the naturalistic respectability of those theories.

On the other hand, some authors defend cluster definitions of MN or physicalism (see, for example, Chalmers 1996; Papineau 2005; Montero 2009; Oppy 2013). They claim that an entity of property is metaphysically natural or physical in virtue of meeting several conditions. For example, Papineau (2005) argues that the physical can be defined as the conjunction of being non-mental and non-vital. As far as I am concerned, cluster definitions generally perform better in their consistency with the agreed cases because additional conditions can always be added to eliminate inconsistencies.

The problem facing cluster definitions is that they are often arbitrary and \textit{ad hoc}. It seems that their various conditions are independent, and it is difficult to see why they should be considered as a single doctrine. What motivates their defenders to add additional conditions into the single doctrine seems to be nothing more than their intention to exclude whatever they do not approve of. Many, myself included, find this subjectivity and arbitrariness unsatisfactory because MN (and/or physicalism) is expected to be a non-arbitrary, systematic, and substantive metaphysical doctrine or framework (see, for example, Stoljar 2010, p. 88).
It is noteworthy that sometimes having a cluster definition cannot be said to be unjustified because it is sometimes sufficient to serve its defenders’ purposes in some particular philosophical debates, such as capturing the characteristics of a typical atheistic and non-mystical worldview for the use of inquiries in philosophy of religion (see, for example, Oppy 2013). However, this is unsatisfactory if we aim not only at those debates, but also at the construction of a systematic and substantive metaphysical doctrine or framework.

Among the authors who defend cluster definitions, Barbara Montero bites the bullet and admits that the above complaint to cluster definitions is correct (2009, p. 181). But she still considers the move to be justified because she holds a pessimistic belief that physicalism cannot be regimented in any better way. Nevertheless, if a non-arbitrary, systematic, and substantive characterisation can be successfully established, which is the aim of this section, there will no longer be any need to fall into the grip of pessimism.

2.4. Problems facing DC

We have seen the strengths of DC. Nonetheless, despite its strengths and popularity among philosophers, the approach has been widely criticised for its several significant shortcomings. Typical criticisms include its inability to solve (1) Hempel’s dilemma, (2) the problem of the naturalistically respectable in history, (3) the problem of non-actualised properties and worlds containing them, (4) the problem of the scientifically inaccessible, and (5) the problem of naturalistically unacceptable scientific theories.

2.4.1. Hempel’s dilemma

Hempel’s dilemma is a seminal problem facing DC. It argues that if DC defines physicalism with reference to physical theories, those theories have to be either those in current physics or those in ideal physics, and both those horns problematic. A seminal formulation of the dilemma is provided by Carl Hempel (1969, 1980), after whom the problem is named (see also Crane and Mellor 1990; van Fraassen 2002). In Hempel’s view, defining physicalism with reference to physical theories is ‘inherently obscure’ (1980, p. 194). The reason is as follows.

If physicalism is defined in terms of current physics, then it is almost guaranteed to be incomplete and false. This is because the fact that current physics is an incomplete description of the world is almost certain. The worry is that any forthcoming theory may prove that physicalism is false.
simply by positing entities and properties that are not posits of current physics (or those constituted by them).

If physicalism is instead defined in terms of an ideal physics, then it is either empty or trivially true. On the one hand, we have no idea what an ideal physics is like, so physicalism will have no content, or be useless. On the other hand, if we provide content to our concept of ideal physics by supposing that it is a set of theories that has successfully described everything (in the actual world), then physicalism is trivially true because everything (in the actual world) is counted as physical. But physicalism is not supposed to be trivially true.

2.4.2. The problem of the naturally respectable in history

Another problem facing DC is whether historical ideas (or figures) can possibly be categorised as naturally respectable. In the history of ideas, many ideas (and theorists who believe in them) are typically considered naturally respectable. However, it seems that there is a dilemma like Hempel’s dilemma facing DC when we take them into account. The dilemma is that we have to either consider most of them as naturally unacceptable, or consider many that are presumptively naturally unacceptable to be naturally respectable instead. The reason is as follows.

One horn is this: if we define MN with reference to current or ideal science, then most presumptively naturally respectable historical ideas have to be counted as naturally unacceptable. This is because few of them (if any) posit only entities and properties that are posits of current or ideal science (or are constituted by them). Hobbesian physics, for example, considers matters to be something like billiard balls in space, but this view has now been proven false by wave-particle duality. Unless the view will be revived in ideal science, DC cannot consider Hobbesian physics to be naturally respectable, but such a revival seems to be extremely unlikely. As Montero writes, ‘It seems to imply that neither Hobbes nor Le Merttrie will be counted as physicalists’ (2009, p. 178).

The other horn is this: if we define MN with reference to every historical idea that is scientific (in the sense of using the scientific method, being respected by scientists as a scientific theory, and so forth), then it seems to be overly inclusive. This is because many historical scientific theories are presumptively naturally unacceptable. For example, vitalism had been a serious scientific theory held by many scientists until the early 20th century, but it is presumptively naturally unacceptable. While it seems difficult to discredit its status as a scientific theory, on this horn of the
2.4.3. The problem of non-actualised properties and worlds containing them

The third problem facing DC is, again, a worry about the narrowness of the inclusive range of it. It is about whether the so-called ‘alien properties’ (i.e. properties not actually instantiated) and worlds containing them can be counted as metaphysically natural. The worry is that, simply because our actual science cannot study other possible worlds and matters within, (almost) every alien property cannot possibly be metaphysically natural, and MN will turn out to be false in (almost) every possible world that contains alien properties. While our standard view suggests that some alien properties are metaphysically natural and that metaphysically naturalism is true in some worlds containing them, DC has difficulties in allowing this to be the case if our actual sciences are referred to exclusively.

Daniel Stoljar (2010) raises the problem as a dilemma. On his view, physicalism can be defined with reference to either a physical theory that is true in the actual world or one that is true in some or other possible world. The former is called ‘actualist theory physicalism’; and the latter is called ‘possibilist theory physicalism’ (p. 76). Actualist theory physicalism is false in worlds in which physicalism is supposed to be true. Suppose that there is a world with a twin-earth with entities and properties such as twin-mass, twin-charge, twin-water, and so on. These entities and properties are almost indistinguishable from their actual counterparts; instead, they are realised by alien properties. Our standard view suggests that physicalism is true in the world, but actualist theory physicalism cannot make this right. On the other hand, possibilist theory physicalism is true in worlds in which physicalism is supposed to be false. Physicalism is presumptively false in, say, worlds with souls and irreducible mental properties, but possibilist theory physicalism cannot make this right because those entities and properties might be posited by physical theories that are true in their worlds.

---

7 I add the bracketed word ‘(almost)’ in all the cases mentioned because there are in fact exceptions, which are that, because of mistakes, instrumentalist reason, and so forth, some properties are posited by our scientific theories even though they only exist in other worlds but not ours. Can this help to avoid the problem of alien properties? The answer can only be no. On the one hand, the categorisation of these posited alien properties as metaphysically natural is arbitrary, and our standard view seems to suggest that there are many natural alien properties that are not metaphysically natural simply because of an arbitrary reason. On the other hand, we can expect that among alien properties that are presumptively metaphysically natural, these alien properties that are posited in this way only make up an extremely small proportion of them.
2.4.4. The problem of the scientifically inaccessible

The fourth problem facing DC is a worry that there might be presumptively natural entities of properties that science cannot access because of, say, some sort of causal isolation, our evolutionary limitation, and so on. If there is a kind of thing that cannot be accessed by science, DC cannot count it as metaphysically natural. As Montero puts it, ‘Do we really want to count, say, a causally isolated collapsed star as non-physical merely because physics could never touch it? I think that the answer is “no”’ (2009, p. 183).

The case of a causally isolated collapsed star might not be fatal to DC because the star might be constituted by those that are scientifically accessible. But it also seems unsurprising that there may be a kind of fundamental particle that is causal isolated, while only having fairly ‘ordinary’ features that, if discovered, would not bother any metaphysical naturalist. In any case, defining the metaphysically natural and non-natural solely with reference to the posits of our knowledge seems to be a mistake because our capability for knowing things might have some limitations that are irrelevant to this distinction.

2.4.5. The problem of naturalistically unacceptable scientific theories

The fifth problem facing DC is the diversity in scientific theories. There are a few current scientific theories that are presumptively naturalistically unacceptable, but if MN is defined with reference to current scientific theories, then they will be by definition naturalistically respectable, which seems to be a mistake. As W. V. O. Quine notes in his criticism of MN, ‘Quantum mechanics today, indeed, in its neoclassical or Copenhagen interpretation, has a distinctly mentalistic ring’ (1995, p. 257). The Copenhagen interpretation posits the concept of the observer effect, according to which some results of a measurement of a quantum process are determined by the way it is measured. The mentalistic interpretation of the Copenhagen interpretation suggests that the observer effect has to be explained by the human consciousness of the observer. If the mentalistic interpretation turns out to be a legitimate scientific theory, then current science seems to be positing some mental properties at the quantum level of the world. This is in conflict with our standard use and understanding of MN, in which any fundamental mental property is counted as incompatible with the doctrine.

No doubt, a metaphysical naturalist can respond that the mentalistic interpretation cannot be categorised as a scientific theory because it violates some principles of the scientific method. Evan Fales (2013), for example, offers a set of arguments for his view that any appeal to disembodied
minds is pseudoscientific, and something similar might be said about mental properties at the quantum level of the world. But there is no guarantee that this kind of move will be successful. In addition, it does not seem appropriate for a definition of MN to rely on this kind of argument. At the very least, according to the principle of having fewer uncommon or idiosyncratic presuppositions, this kind of argument should be avoided.

2.5. The similarity approach (SA)

Despite the fact that many revised definitions of MN and physicalism have been proposed in the literature in response to the problems above, some suggest that a simple modification of DC is sufficient to solve these problems. The suggestion is to allow extensions of current science, constrained by adequate similarity to existing science (see, for example, Lewis 1983; Ravenscroft 1997; Braddon-Mitchell & Jackson 2007). I call this approach the similarity approach (SA).

There are two conceptions of similarity that proponents of SA have to choose between: qualitative similarity and theoretical similarity. The former is the similarity between properties or entities; and the latter is the similarity between the theoretical descriptions of the properties or entities under consideration. Lewis makes use of the former when he defines materialism as the doctrine that ‘something not too different from present-day physics, though presumptively somewhat improved, is a comprehensive theory of the world’ (1983, p. 361). Ian Ravenscroft makes use of the latter when he defines physical properties as ‘those properties sufficiently similar to the properties quantified over by present-day physics’ (1997, p. 424). David Braddon-Mitchell and Frank Jackson are going both ways when they define the physical as ‘those roughly like those in science or a conservative extension of it’ (2007, p. 34).

Although I consider these choices to be important to our concerns, I do not wish to address them or take any stance here. In order to address them, we have to engage with some general problems in theories of similarity (or resemblance), which is an independent area of philosophy being intensively researched, and which should be independently addressed.

---

8 It is noteworthy that Lewis is not defining physicalism merely by SA, which is the approach to defining the physical with reference to those that are similar to current physics. This is shown by his line, ‘though presumably somewhat improved, is a comprehensive theory of the world’. What he has in mind is an intermediate position between defining the physical with reference to current physics and defining the physical with reference to ideal physics. The idea is to define the physical with reference to ideal physics, but only when ideal physics is similar to current physics. With this in mind, most problems facing DC other than Hempel’s dilemma, such as the problem of the naturalistically respectable in history and the others, are presumptively left unaddressed by Lewis, because the properties or entities under consideration are presumptively not posited by ideal physics. Nevertheless, Lewis’s usage of SA can be taken as a useful reference for our purpose.
Setting general problems in theories of similarity aside, I consider SA to be an attractive approach. It can lead to a reasonable extension of the inclusive range of DC, and can therefore avoid the narrowness of the first horn of Hempel’s dilemma (which is to refer to current scientific theories only). Future and historical scientific theories can be counted as naturalistically respectable through being adequately similar to current science. Alien properties and scientifically inaccessible properties can also be accounted for if any of them share this similarity. It is less clear how the approach can account for naturalistically unacceptable scientific theories because MN is defined with reference to similarity to current science, which by definition includes all current scientific theories, presumptively naturalistically respectable or not. But I believe that the proponent of SA can respond by providing a slightly revised version of SA which suggests while there is a similarity between most current scientific theories, there are a few exceptional cases that do not share this similarity, and we therefore should not count them as naturalistically respectable.

2.5.1. SA and alien naturalism

A possible objection to SA is that it cannot really solve the problem of non-actualised properties and worlds containing them. I shall discuss the objection in detail even though I am not a proponent of SA, because the objection and my reply apply equally to the approach I am going to propose. According to the objection, SA is modally relative, in the sense that it makes an assumption about which possible world the actual world is and thus cannot account for counterfactual judgments made within worlds containing alien properties. The reason is as follows. SA defines the metaphysically natural with reference to our actual science, but otherworld inhabitants might have developed their otherworld sciences (such as a science of soul) in terms of the alien entities or properties in their worlds. As a result, their judgments about whether a thing is metaphysically natural or not must be different from ours. A similar point can be made about future generations. This is because future sciences will probably be different from current science, and hence SA has a difficulty in explaining how judgments about metaphysical naturalness can be made by future generations (or our future selves). This may be problematic for some.

I do not find the objection to be compelling, because I disagree that the above stories about otherworld inhabitants and future generations are problematic for SA. An assumption of the objection is that MN has to be modally and chronologically neutral – in the sense that it makes no assumption about which possible world the actual world is and when the present is – and be able to account for counterfactual and future judgments. It is difficult to see why the assumption has to be made. There
are, indeed, some concepts that are modally and chronologically neutral, such as the concept of a triangle, but many others are modally and chronologically relative, and we nonetheless find them useful and have no reason to eliminate them. Consider natural kind terms like ‘water’. As we all have learned from Hilary Putnam and Saul Kripke, the meaning of the term ‘water’ depends on the empirically discovered fact that the watery stuff in the actual world is H₂O, but this is not true for otherworld inhabitants, nor is it the way our ancestors understood the term. In addition, it is not impossible that future generations will discover that our chemistry is, in fact, mistaken and the watery stuff in the actual world is not H₂O but something else, and hence understand the term ‘water’ differently. None of these facts show that the meaning of the natural kind term ‘water’ is in any sense objectionable. Taking this into account, it is difficult to see why we should demand modal and chronological neutrality from a characterisation of MN.

2.5.2. The dilemma of the right kind of similarity

What I consider to be the fatal failing of SA is what I call the dilemma of the right kind of similarity. The concern is as follows. To have a level of similarity, precise or not, that acts as a boundary between the metaphysically natural and non-natural, we need to have the right kind of similarity to act as the measure. Available options include (1) overall similarity and (2) similarity in a certain respect. However, on the one hand, (1) cannot work, as any boundary drawn by using overall similarity will not correspond to the boundary we expect for MN because some presumptively natural things are obviously less similar to current science than some presumptively non-natural things in terms of overall similarity. I will discuss this in detail in the coming paragraphs. On the other hand, we might, instead, follow (2) and consider all presumptively natural things to be similar in some specific respect, such as being spatiotemporal. Unfortunately, in this case the metaphysically natural is, in fact, defined in terms of those specific respects, not really in terms of similarity. Hence, if we accept this option, we are, in fact, accepting some other approach to characterising MN and thus giving up SA. In addition, we have seen that other such approaches rarely work. Different from the problems with the general use of the concept of similarity I mentioned earlier, the dilemma of the right kind of similarity is particular to SA and cannot be avoided, even if we take the plausibility of the common usage of the notion of similarity for granted.

The matter of concern is then whether some presumptively natural things are less obviously similar to the posits of current science than some presumptively non-natural things in terms of overall similarity. I think there is an obvious example of such a thing. Compare the posits of Hobbesian
physics, string theory, and David Chalmers’s panpsychism. The dissimilarity between strings and Hobbesian matter is tremendous. Strings have a fairly unique nature. For example, it is suggested that they act in eleven dimensions, and that they are not obviously spatiotemporal, and so forth. They are tremendously different from what Hobbesian physics posits – physical particles that are like billiard balls, only acquiring a new motion when another is taking, and thereby pushing it away from, its current spatiotemporal location. Meanwhile, Chalmersian panpsychism (or as many philosophers of mind call it, Russelian panpsychism) is consistent with all current physical theories, such as string theory. What it adds to the posits of those theories is merely a nature intrinsic to them, which is phenomenal (Chalmers 1996, pp. 153-154). A material object described by the doctrine in combination with current physics seems to be very similar to one that is solely described by current physics. Therefore, it is hard to see how there could be a certain level of overall similarity of entities or properties that MN can use to count Hobbesian physics as naturalistically respectable while counting Chalmersian panpsychism as naturalistically unacceptable.

Some might object that even though material objects described by Chalmersian panpsychism seem to be similar to those of current physics, the so-called phenomenal intrinsic nature, as a property, is not similar to any property posited by current science. But I believe that a careful examination of Chalmersian panpsychism and current theories of properties seems to show that this may not be the case. It is important to note that the fundamental phenomenal properties that Chalmersian panpsychism posits are not phenomenal in the sense our phenomenal experiences are. They are, as Chalmers sometimes calls them, protophenomenal properties, by which our phenomenal experiences are constituted (127, 154). In this case, protophenomenal properties are not really as complex as our phenomenal experiences but are simple and ‘thin’. Further inquiry into how these protophenomenal properties can fit into our metaphysical framework is a matter of concern in theories of properties, which I cannot discuss in detail here. But the basic (and rough) idea is that some dispositional or causal properties described in science are roles played by protophenomenal properties, or have a non-dispositional, non-causal, and ‘qualitative’ (as some theorists of properties call it) side that is protophenomenal. If current science has posited any such role-player property or qualitative side of a property, such as the spatial nature of an object, then protophenomenal properties can be similar to it because they are simple and ‘thin’ enough to not be significantly different from their counterparts.
that are not protophenomenal. \(^9\) It is difficult to see why their similarity is any less than that between the properties of strings and those of Hobbesian matter.

In the face of this problem, those who are sympathetic to SA have two options other than repudiating it. The first option is to accept that the inclusive range of MN is different from what most believe. For example, one may agree with Chalmers that Chalmersian panpsychism is, indeed, naturalistically respectable (1996, p. 170). Though this option is attractive, it is costly because many of our assumptions concerning the metaphysical naturalness or non-naturalness of things (and the naturalistic respectability of theories) might have to be revised. For example, other than counting Chalmersian panpsychism as naturalistically respectable, we might also have to count Hobbesian physics and Newtonian physics as naturalistically unacceptable, to count pantheism as naturalistically respectable, and so forth. With this in mind, many discourses concerning MN in philosophy, for or against the doctrine, have to be considered as wrongheaded. This consequence, I suppose, may be too costly even for many who are sympathetic to Chalmers.

The second option is to keep our assumptions about MN and develop or revise SA to bring forth a regimentation of its inclusive range which could be consistent with our assumptions about the naturalness of things. My option is an intermediate proposal between repudiating SA and the second option. More precisely, I will develop an original approach to DC with the same strategy that SA uses, whereby extensions of current science are allowed.

Braddon-Mitchell and Jackson, two proponents of SA, rightly notice and point out that the use of similarity in SA invites vagueness (even though it is not clear whether they have the dilemma of the right kind of similarity in mind). They thus write, ‘The vagueness in the definition can perhaps be left to advances in philosophy and indeed in science itself to sort out’ (2007, pp. 34-35). Although the characterisation of MN I will develop and defend is a radically revised proposal, I still consider it to

\(^9\) Some theorists of properties, such as Armstrong (1961) and Simon Blackburn (1990), believe that science in principle studies and describes only dispositions but none of their role-players or qualitative sides. This view might challenge what I have been arguing here, because any non-dispositional properties or qualitative sides of dispositional properties will, accordingly, be tremendously different from what science posits. A matter of concern here is that the view is fairly contestable and controversial (see, for example, Smart 1963; Ellis 2010; Ney 2015), but I believe that even if it is correct, we should carefully distinguish what the scientist actually says and how the metaphysician interprets her words. While this view belongs to the latter, DC should only take the former into consideration. The spatial nature of an object is an example. The metaphysician may believe that descriptions of the spatial nature of objects in science are simply about where dispositions factually or counterfactually manifest, and thus is, ultimately, dispositional in nature. Even if this is correct, the scientist merely posits the spatial nature of objects; and it is the metaphysician’s work that interprets this in terms of metaphysics and reduces it to dispositional descriptions. Hence, even if metaphysicists such as Armstrong and Blackburn are correct, SA still allows for protophenomenal properties to bear similarities to posits of current science that are yet to be interpreted by the metaphysician and thus be possibly metaphysically natural.
be a development of SA because they share the same strategy – which is to search for an extension of current science. With this in mind, a certain level of vagueness will unavoidably persist. But I believe that my approach can provide the extension of the inclusive range of MN from current science that SA aims at, and at the same time can avoid the dilemma of the right kind of similarity. Hence, I hope that my proposal is at least a step forward in the sense they expect (if it is not to be considered as a rival but a development).

2.6. The historical path approach (HPA)

I call my approach the historical path approach (HPA). HPA is based on a concept I use to replace the similarity relation used by SA, which is what I call the historical path (HP) of modern science (and in the case of physicalism, HP of modern physics). HP is the process of developing and modifying scientific concepts over the course of history, along which they are modified and finally developed into the versions in modern natural science. What I call ‘concepts’ include theoretical contents, methodologies, measurements, and so on. Examples include the concept of momentum within Newtonian physics, the concept of positron in current particle physics, the physicist’s preference for mathematically quantifiable and simple theoretical posits, etc.

What is suggested here assumes a distinction between concepts and theories, with the latter including Newtonian physics, the theory of relativity, string theory, and so on; and HP is about concepts, not theories. Not every philosopher will agree with the distinction because these concepts themselves might be, as some theories of concepts suggest, small theories. I take no stance on this controversy; and I suggest that those who disagree with the distinction can replace the term ‘concept’ here with small theories that are components of big theories or theoretical frameworks. Some might wonder why I intend HP to be about concepts but not theories. The motivation for my decision is my belief that only when HP is about concepts can it resist the problem of overinclusion, which is the worry that HP will be overly inclusive and thus useless. I will return to the problem and address it in detail in Section 2.6.3.

Let us return to HP. Even though the concepts within Newtonian physics might be very different from those within string theory, they have some historical connections. For example, the concept of gravity in Einsteinian theory of relativity is developed from its counterpart in Newtonian physics, and it, in turn, has developed into the concept of quantum gravity in string theory. Of course, a lot of scientific concepts do not have older counterparts. For example, it is less clear that the concepts of positron or Higgs boson have older counterparts. But it can be said to have been developed in close
relation to concepts we already have, such as those of electron or photon, or that of physical particle in general. Historical connections as such constitute HP. On the other hand, some theoretical contents and methodologies in other scientific theories – such as Aristotelian teleology and some versions of vitalism – can be said to be relatively more disconnected from HP, even though these theories and the concepts within might have their rational basis in their time.\(^\text{10}\) This is because many of the central concepts within were wholly refuted rather than modified and absorbed by new theories in a historical progression. For example, the concept of telos has been abandoned rather than modified in physics. To put this simply: many of their central concepts fail to have descendants in current science.

If my hypothesis about the existence of HP is correct, then a definition by the historical path (DHP) can be formulated as a definition of MN. The basic formulation is as follows:

**DHP:** Every entity or property instantiated in the actual world is metaphysically natural – in the sense of being posited by some scientific concept(s) that is on HP.\(^\text{11}\)

It is noteworthy that not all concepts posit entities or properties. For example, some methodological concepts may not posit anything. What is important is that if an entity or property is metaphysically natural, then there should be at least one concept that posits it.

**2.6.1. Future science**

The basic formulation above cannot allow future scientific theories to be naturalistically respectable, because they are very likely to have new concepts that are neither current concepts nor ancestors of them. But we a modified version of DHP can solve this problem:

**DHP2:** Every entity or property instantiated in the actual world is metaphysically natural – in the sense of (a) being posited by some scientific concept(s) that is on HP or (b) being likely to appear on HP in the future.

Condition (b) is added to the definition as a sufficient but not necessary condition alternative to condition (a). We may have some reasonable expectations about whether a development of a concept

\(^{10}\) I am here setting aside the question in philosophy of science of whether these views can be legitimately called ‘science’.

\(^{11}\) I did not add the clause of the original version of DC, ‘or being exclusively constituted by (i.e. reducible to or realised by) those posits’, into DHP. It should be added in, but since I am going to develop some fairly complex variants of DHP, I purposefully set it aside to avoid unnecessary confusion.
through scientific advancement is more or less likely, such as the expectation that we may discover a physical particle that quarks are composed of. With the assumption that such expectations are possible, one can reasonably tell, for example, that a paradigm shift in physics to some New Age theories that posit a psychic universe is highly unlikely, and that discovering ghosts in neuroscience is highly unlikely. Some might wonder whether the expectation here is a matter of similarity to current theories and hence a retreat to SA. I think no; at least it is not merely a matter of similarity. Many other considerations can be involved to bring forth a more sophisticated expectation than what mere similarity can allow for. Examples include the scientific method in general (which includes principles such as the need for empirical observations or experiments, the preference for repeatability, and so on), theoretical virtues in general (such as simplicity, explanatory width, and so on), methodologies used by experts in different branches of science, tendencies of what new discoveries turn out to be like, technological limitations, and so on.

Admittedly, these expectations still have to be fairly imprecise because they are often not based on very precise technical details, but rather some very abstract conceptions of what current science is like. To be more precise, they often have to be made from the perspective of a sufficiently informed layman, not from the kind of expectation a scientist would have with regard to the forthcoming discoveries based on current findings. This is because if the expectations are limited to precise technical expectations, such as the expectation of a nanophysicist working on silver that silver might be used as a conductor for such and such a purpose, then revolutions in science cannot be accounted for. We want something less narrow to do the job. With this in mind, the kind of expectation needed might be like what most reasonable laymen who are sufficiently informed about current science would expect, say, science to be like after a hundred years.

Certainly, some would be unsatisfied with the imprecision involved. But it is difficult to see how it could be a real failing of the approach. As Ravenscroft argues in his defence of SA, on the one hand, what is most important is to be able to categorise most paradigmatic cases of the non-physical; on the other hand, our conception of the physical is indeed imprecise and vague; and thus if a characterisation of the physical is committed to vagueness, the commitment might simply be an explanation of the vagueness within our conception of the physical, rather than a failing (1997, pp.

---

12 No doubt, a scientific concept is likely to appear on HP in the future because of social or psychological factors that are pretty indifferent to genuine theoretical considerations, such as corruption in the scientific community, political upheaval, religious influence, destruction of the human civilisation, and so forth, which most of us would consider to be irrelevant to the discussion here. By the clause ‘are likely to appear on HP in the future’, I only mean those that are likely to appear because of unbiased and undisrupted scientific inquiries, which are not intervened by the above factors.
I believe that the same can be said about DHP2. On the one hand, it is fairly obvious that the kind of expectation DHP2 makes use of enables us to categorise most paradigmatic cases, such as God, the fundamentally mental, undiscovered particles, and so forth, as we expect of MN. None of them seems to fall in the marginal area between the metaphysically natural and non-natural where vagueness is an issue. On the other hand, the metaphysical naturalness (or naturalistic respectability) of some cases, such as intentionality, is indeed difficult to determine; thus, the vagueness of the boundaries DHP2 is committed to may be helpful in explaining why such cases happen. If the reader is unconvinced, she can think of the cases of colours or morals, in which we can typically recognise what is, say, red or morally evil, even if there are borderline cases that we cannot decide; and it is not obvious that any of our characterisations of redness and moral evil have to be considered to be a failure if there is vagueness at the boundary.

It is noteworthy that it is not impossible for there to be a very different HP from the perspective of future science. For example, it is possible for a return to Aristotelian teleology to occur. A future path as such might count our current important scientific concepts to be ones that fail to have descendants. My view on this is exactly like my argument earlier on in Section 2.5.1 that SA is not defeated by the same concern. It is not obvious that the concept of MN is supposed to be modally and chronologically neutral and be able to account for future judgments.

Before I further develop my approach, two problems must be addressed here. They are (1) the problem of disconnection and (2) the problem of overinclusion. The two problems are not consistent with each other. The former suggests that HP does not have a substantial length because scientific theories are, by and large, independently developed but not developed through modifications of concepts of older scientific theories. The latter suggests that HP is inclusive of too much, which includes concepts that are presumptively naturalistically unacceptable. Only when these two problems are addressed can the nature of HP (and HPA) be made clear, which will allow us to continue to see how the approach can be developed.

Through addressing these two problems, we can also see a large part of how HPA can resist the worry motivating the dilemma of the right kind of similarity. The worry is there in the case of SA because no level of overall similarity can act as the boundary that allows us to count things as metaphysically natural or non-natural as we expect of MN. Addressing the two problems allows us to see how the inclusive range of HP is shaped and can be used by MN as a boundary, which I believe to be superior to the boundary drawn by SA.
2.6.2. The problem of disconnection

Let us begin with the problem of disconnection. It is a scepticism about the existence of any HP that has a substantial length. Some friends of Thomas Kuhn’s thesis of paradigm shifts might be particularly drawn to this view. They claim that major developments in the history of science occur when paradigm shifts (i.e. radical changes of the most fundamental assumptions within short periods of time) happen. With these radical changes, it is unclear that a path can really be drawn because there might be many disconnections between scientific theories. They might make use of the following idea of Kuhn:

The transition from a paradigm in crisis to a new one from which a new tradition of normal science can emerge is far from a cumulative process, one achieved by an articulation or extension of the old paradigm. Rather it is a reconstruction of the field from new fundamentals, a reconstruction that changes some of the field’s most elementary theoretical generalizations as well as many of its paradigm methods and applications. During the transition period there will be a large but never complete overlap between the problems that can be solved by the old and by the new paradigm. But there will also be a decisive difference in the modes of solution. When the transition is complete, the profession will have changed its view of the field, its methods, and its goals. (1996, pp. 84-85)

It should be noted that it is not very obvious that Kuhn himself would oppose my idea of HP since the cited passage above is not written for the discussion going on here. Although he is discussing the historical progression of science like I am, he is engaging in a discussion of how scientists actually develop theories and solve problems in history, not of whether there is an abstract historical lineage with a substantial length that can be used to define MN or physicalism. Nevertheless, it remains possible for sceptics about HP to make use of this view to develop an objection.

A deep debate about philosophy and history of science (and Kuhn’s views) would be outside the scope of this dissertation, but my view about this can be stated here very briefly. In my view, it is difficult to see why the sceptic is correct. Even if Kuhn’s theory of paradigm shift is correct, the case is only that there is no HP in a narrow sense, in which concepts such as <gravity> and <space> are defined or understood in exactly the same ways. However, this does not imply that there cannot be some kind of HP in a broader sense, on which old concepts are found to be modified and absorbed by the new theories. In fact, Kuhn himself believes there are ‘borrowed elements’ from the old paradigm,
‘both conceptual and manipulative’, that are incorporated into the new ones (149). The examples he provides include <planet>, <space>, and so on (128, 149). It is true that he believes the borrowed elements are more radically revised than many believe. In his view, they are not employed in the traditional ways as they were in the old paradigm, because they form new relationships with other concepts in the new paradigm. But it is hard to see why this alone is in conflict with the idea of HP.

2.6.3. The problem of overinclusion

My response to the problem of disconnection shows how HP can extend over eras. The other problem that follows is the problem of overinclusion. Some might, contrary to the friends of the problem of disconnection, consider HP to be overly inclusive and thus useless because it is possible to understand our current science as descending from belief systems that appear to be mostly naturalistically unacceptable. For example, it can be suggested that modern chemistry is a descendant of alchemy, which is full of magical concepts.

To respond to this objection, two conceptions have to be distinguished. Recall that HP is about concepts (or small theories), not belief systems (or big theories). For example, HP is not about whether alchemy as a belief system is the ancestor of modern chemistry but, rather, which particular concepts in alchemy are the ancestors of those in modern chemistry. In fact, as I have mentioned earlier, the problem of overinclusion is why I intend HP to be about concepts – because an HP about theoretical frameworks, belief systems, or big theories cannot avoid being overly inclusive, but an HP about concepts can. Alchemy, as a discipline, is a grand belief system containing many concepts, including its theoretical concepts about magic, classifications of chemicals, measurements, and so on. As history progressed, some of them have been developed and absorbed by chemistry while others have been given up entirely. HPA has no problem taking some concepts in alchemy, such as its observations of chemical reactions, to be ones on the path, thereby taking them to be naturalistically respectable. Yet it notes that the theoretical concepts about magic within alchemy have been entirely abandoned and should therefore be considered naturalistically unacceptable. To put this simply: only some but not all of the concepts within alchemy are on the path, and none of those on the path seem to be presumptively naturalistically unacceptable.

With the problem of disconnection and the problem of overinclusion solved, and the nature of HP made clear, the worry motivating the dilemma of the right kind of similarity, by and large, dissolves. For cases such as Hobbesian physics, even though they are obviously not as similar to current science as Chalmersian panpsychism is, most if not all of their main concepts seem to have
been continuously modified and finally developed into those in current science. On the other hand, the concept of proto-consciousness in Chalmersian panpsychism, as far as I am concerned, simply is not on HP, nor is it what most of us consider to be likely to appear on the path. Certainly, it can still be questioned whether HP can really exhaustively cover every entity or property we take to be metaphysically natural. The possibility that there are some other counterexamples cannot possibly be expunged, even though, as far as I am concerned, there are no such examples. Nonetheless, the problem is, I believe, significantly less obvious than it was in the case of SA.

2.7. The alien and the scientifically inaccessible

No doubt, DHP2 cannot solve the problem of non-actualised properties and worlds containing them and the problem of the scientifically inaccessible, because alien properties and scientifically inaccessible properties are not things we can discover in future science. But a slight modification of the definition can solve the two problems:

**DHP3:** Every entity or property instantiated in the actual world is metaphysically natural – (1) in the sense of (a) being posited by scientific concepts that are on HP or (b) being likely to appear on HP in the future, or (2) in the sense of playing only a set of theoretical roles that is closely similar to one played by a metaphysically natural entity or property in the former sense.\(^{13}\)

Condition (2) is added to the definition as a sufficient but not necessary condition alternative to condition (1). What is meant by a ‘theoretical role’? It means, very roughly and abstractly speaking, an attribute that a theoretical framework like Einstein’s theory of relativity offers to describe and/or interpret the characteristics of an entity or property. Examples include the possession of a mass, charge or size property, being constituted by some more fundamental entities, an identity as an evolved creature, and so on, which are offered by theoretical frameworks such as Einstein’s theory of relativity, string theory, Darwinian theory of evolution, the current paradigm of neuroscience, and so forth. It is noteworthy that here I assume a distinction between a property and an attribute. While the former is a real constituent of the world, the latter is conceptual and descriptive.

---

\(^{13}\) Of course, only scenarios in which the entity or property under consideration is correctly described by the theory are taken into account. I set aside what it means to be correctly described by the theory, since this will lead us into matters of theories of reference and theories of truth, which are not our concern here. What has been said is, I believe, sufficient for our purposes.
No doubt, what I am describing here is unspecific and highly abstract, as well as relying on examples rather than a definition. Hence, a lot more work has to be done with respect to what exactly a theoretical role is and what kinds of theoretical roles there are. In fact, these matters are purposefully left undetermined because a study of the history and philosophy of science as such is beyond the scope of this dissertation. In addition, they should be left undetermined in order to allow readers with different views on these matters to fill in their own preferred views and thereby find the definition useful.

Needless to say, the close similarity between sets of theoretical roles is taken into account, and hence the use of similarity is reintroduced. But this does not mean that DHP3 faces the dilemma of the right kind of similarity like SA did. This is because what is required here is much more specific and thus demanding than the notion of overall similarity used by SA. More precisely, two sets can be considered as closely similar when almost all of their crucial roles have counterparts and when the crucial roles are closely similar to their counterparts. With this in mind, the close similarity between sets of theoretical roles are not intended to allow, say, those of the posits of Hobbesian physics and of string theory to be said to be similar. Therefore, the very broad and undemanding overall similarity in SA that is intended to connect all naturalistically respectable theories – such as Newtonian physics and string theory – together while being exclusive of all naturalistically unacceptable theories is not involved, and thus the worry that motivates the dilemma of the right kind of similarity is not facing DHP3.

Let me provide some examples. Consider an alien or scientifically inaccessible particle $x$ whose nature is exactly like that of an electron, with the exception that it has a different mass property. It can be counted as metaphysically natural as we expect because almost all of its nature is exhausted by the theoretical roles played by an electron, including its possession of a mass property, with the exception of a quantitative difference with respect to that property.

Meanwhile, it is not obvious that condition (2) will allow presumptively metaphysically non-natural things to be counted otherwise. Consider our earlier example, Chalmersian panpsychism. If a particle $y$ shares the nature of a quark but, in addition, has an intrinsic protophenomenal nature as posited by the doctrine, it will not be counted as metaphysically natural by condition (2). The reason is that no theoretical role played by metaphysically natural entities or properties can be played by its intrinsic protophenomenal nature. Consider also a ghost that is a fundamental entity. Of course, it can play a lot of theoretical roles that are similar to the ones that are posited by, say, theories in animal behaviour studies. In addition, it may possess no property other than those that play those roles, since
it is fundamental. Can it therefore be described as playing only a set of theoretical roles that is closely similar to one possessed by an animal because all of its properties are shared by an animal? The answer is no, because it lacks the roles of being an evolved creature, being an embodied or constituted cognitive system, and so on, which are important members of the set of theoretical roles that is played by an animal studied by animal behaviour studies.

For the same reason, condition (2) cannot replace condition (1). To be more precise, much like the case of a ghost, posits of Newtonian physics, for example, cannot be considered as metaphysically natural by condition (2). The reason is as follows. Even though many contemporary physicists do not consider Newtonian laws to be ‘false’, those laws are, at best, considered as approximately true values within some specific circumstances. However, the posits of Newtonian physics strictly follow the laws. For example, their mass properties are not convertible to energies. Hence, they have sets of theoretical roles that are not similar to any posit of current science, and thus have to be counted as metaphysically natural by condition (1) instead of condition (2).

Another matter of concern is whether DHP3, by being a disjunction of two sufficient conditions of the metaphysically natural, will fall prey to the objection against cluster definitions I provided earlier in Section 2.3. The worry is that putting irrelevant conditions together is subjective, arbitrary, and \textit{ad hoc}. I admit that DHP3 is, to a certain degree, \textit{ad hoc}. This is because the second condition is added solely to account for the problem of non-actualised properties and worlds containing them and the problem of the scientifically inaccessible, and there would be no need for it if DHP2 were not facing the two problems. Nevertheless, it does not seem to be problematic, since condition (2) is a natural extension of condition (1) to deal with special cases. On the contrary, the cluster definitions I criticised earlier are ones whose conditions lack substantive connections and are put together subjectively and arbitrarily with the sole purpose of excluding what their proponents disapprove of.

\textbf{2.8. Naturalistically unacceptable scientific theories}

We have seen how HPA can survive objections from Hempel’s dilemma, the problem of the naturalistically respectable in history, the problem of alien properties and worlds containing them, and the problem of the scientifically inaccessible. What is left is the problem of naturalistically unacceptable scientific theories. In response to this problem, it can be suggested that there is something like a \textit{sudden divergence from the path}. This is to suggest that theories such as the mentalistic interpretation of the Copenhagen interpretation, by positing concepts such as the role of
consciousness in the mechanics of fundamental physics, are not going along the general direction of gradual changes, which most concepts on the path are moving towards.

The concept of a sudden divergence from the path is not *ad hoc*, because it not only allows us to account for the reason why some scientific theories are presumptively naturalistically unacceptable, but also is useful in its own right, in the sense that it can (partially) explain the research directions of some scientists (and philosophers). For example, some scientists are motivated to propose new interpretations of quantum mechanics that are alternatives to the Copenhagen interpretation, in order to avoid its use of the (possibly mentalistic) observer. Consider the example of theoretical physicist John Cramer, the proposer of the transactional interpretation. He writes:

> The knowledge interpretation’s account of state vector collapse and nonlocality is internally consistent but is regarded by some (including the author) as subjective and intellectually unappealing. It is the source of much of the recent dissatisfaction with the Copenhagen interpretation. The author has proposed an alternative and more objective interpretation of the quantum mechanics formalism called the transactional interpretation. (1988, p. 228)

In addition, some scientists and philosophers particularly expect and welcome such proposals (see, for example, Price 1996). A possible explanation of their motivation offered by HPA is that they might want to avoid a sudden divergence from the path. If this explanation is correct, then there can be a rational basis for this kind of motivation other than a mere emotional discomfort about introducing the fundamentally mental. After all, it might be rational to have a conservative attitude.

### 2.9. The usefulness of HPA

I advocate that HPA is not only attractive and defensible as I argued above, but also useful because it has strengths in explaining some patterns of behaviour of scientists (and philosophers) and the folk.

Consider the way I handle the problem of naturalistically unacceptable scientific theories, in which I appeal to their sudden divergences from the path. I have argued that it not only allows us to account for the reason why some scientific theories are presumptively naturalistically unacceptable, but also is useful in (partially) explaining the research directions of some scientists (and philosophers).

There is an additional reason why I consider HPA to be useful, which is that it can explain the folk intuitions about and usage of the terms ‘naturalism’, ‘natural’, and ‘supernatural’. It seems that the folk often distinguish the natural and the supernatural by assessing whether something is posited
by some adequate extension of current science. For example, it seems that a future world described by a work of science fiction is typically considered as natural, even if it contains entities that are not really posited by science, while magic described by some other fictions is typically considered as supernatural. HPA provides two attractive explanations for the folk intuition. The first is that the folk have an expectation that the entities in the fiction would be posited by future scientific theories. The second is that the folk have an intuition that the entities in the fiction are playing sets of theoretical roles that are similar to those played by posits of current scientific theories or expected future scientific theories.

I have noted in Section 1.4.1 that metaphysical naturalness is not merely a folk concept, since its meaning has been altered when it is used in philosophy. For example, while the folk often count werewolves, aliens, and UFOs to be supernatural, I doubt any philosopher will share these views. Nonetheless, what is attractive about DHA is that it also provides an explanation of the way folk naturalism evolved into the philosopher’s MN. The explanation is as follows. The folk often count werewolves, aliens and UFOs to be supernatural simply because those things are not included in normal science. The philosopher, on the other hand, by her more sophisticated knowledge of the history of science and the theoretical roles of entities, has regimented the boundary between the metaphysical natural and non-natural. Hence, cases such as werewolves, aliens, and UFOs are counted as metaphysically natural.

2.10. Developing HPA into a metaphysical thesis

Up to this point, HPA has been mainly based on historical and sociological considerations. Metaphysics and ontology have not been involved much. It could be suggested that HPA – and perhaps DC in general – is a sociological thesis because the scientific discipline and its history is a sociological issue. This might be, for many, counterintuitive and unsatisfactory because when we are using the term ‘MN’, we are usually trying to express metaphysical concepts, not merely sociohistorical concepts. It might be acceptable that MN has some sociohistorical roots, but intuitions also expect that these theses have substantive metaphysical contents.

In response to this possible complaint, here I briefly suggest two possible ways to understand HPA as, or develop HPA into a metaphysical thesis. I call them the *minimalist thesis* and the *metaphysical paradigm thesis*.
2.10.1. The minimalist thesis

The minimalist thesis is an understanding (or an interpretation) of MN defined by HPA. According to the thesis, MN should be read as a metaphysical doctrine because its exclusive commitment to entities or properties posited by HP-related scientific concepts is a decision in ontological economy, which is itself not an HP-related scientific concept, but a metaphysical commitment. Consider any version of DHP I proposed. All of them begin with the clause, ‘every entity or property instantiated in the actual world is metaphysically natural’. We have seen that the term ‘metaphysically natural’ is defined with reference to HP, but the clause ‘every entity or property instantiated in the actual world is’ is not.

With this in mind, the metaphysical naturalist is a minimalist metaphysician who always restricts her metaphysical commitments to entities and properties posited by HP-related scientific concepts and is unwilling to have any additional commitment. On the contrary, the metaphysical non-naturalist, from the perspective of the minimalist thesis, is one who goes beyond the minimalist restriction and adds additional metaphysical commitments. Consider Oppy’s claim that ‘theism is an extension of naturalism’ (2013, p. 88). In his view, theism has all naturalistic commitments that metaphysical naturalism has, but it adds additional supernatural commitments, such as God, to it. While he does not define MN and non-naturalism by HPA, his view provides us with a good hint on how we can understand the relation between MN and non-naturalism defined by HPA.

Some readers might find the minimalist thesis unsatisfactory. This is because some might not only believe that metaphysical naturalism is a metaphysical doctrine, but also believe that the concept of metaphysical naturalness is a metaphysical one. While the exclusive commitment to HP-related scientific concepts is a metaphysical doctrine, HP-related scientific concepts are still defined solely in sociohistorical terms. I suggest that these readers may, instead, consider the metaphysical paradigm thesis.

2.10.2. The metaphysical paradigm thesis

The metaphysical paradigm thesis is a new definition of MN that can be developed from HPA. The method is to make use of HP to locate a metaphysical paradigm within science. It is as follows. We have seen that HP is a path of concepts (or small theories) within theories (or big theories or theoretical frameworks). Among theories that contain concepts on HP, only some have most if not all of their main concepts on HP. For example, we have seen that alchemy contains concepts on HP, but many of its main concepts are not on HP. On the contrary, Hobbesian physics has most if not all of
its main concepts on HP. Let us call these theories having most if not all of their main concepts on HP the *core theories*.

Locating these core theories allows us to locate the metaphysical paradigm. Most of the core theories are ones in or after the early modern period, such as Hobbesian physics, Newtonian physics, and the theory of relativity; and many of them have important roles in shaping HP. There might be some metaphysical commitments shared by the core theories (or the scientists who are working on them). Presumably, these commitments not only are the commonalities between the core theories, but also explicitly or implicitly served as part of the guiding forces of the shaping of HP. Instead of defining MN with reference to the scientific concepts on HP, HPA may instead define MN as a metaphysical paradigm that has the metaphysical commitments of the core theories as its contents. The definition is as follows:

**DHP4:** Every entity or property instantiated in the actual world is metaphysically natural – in the sense of being compatible with the metaphysical paradigm shared by the core theories.

What are these metaphysical commitments? We might have examples like ‘everything is constituted by (something similar to) theoretical entities in physics’, ‘everything operates according to laws of nature’, ‘there are no fundamental mental properties’, and so on. These are all suggested possibilities and I am not attempting to argue for or against any of them at the moment. My sole attempt here is to show that the metaphysical paradigm thesis is one of the possible ways to develop HPA into a metaphysical thesis.

The metaphysical paradigm thesis probably has multiple metaphysical commitments as the content of MN. One might suspect that this falls into the kind of cluster definition without simplicity and systematicity I criticised earlier in Section 2.3. But this is not true. As a definition of MN, the metaphysical paradigm thesis merely defines MN with reference to the metaphysical paradigm, which is the set of metaphysical commitments as a whole. Each metaphysical commitment is referred to in virtue of being a member of the set, and thus the inclusion of them is not subjective, arbitrary, or *ad hoc*. On the contrary, the kind of cluster definition I criticised earlier refers to individual conditions one by one, which is why it seems to be subjective, arbitrary, and *ad hoc*. Even if, fortunately, they all turn out to be members of the metaphysical paradigm that the metaphysical paradigm thesis spells out, they are not referred to in virtue of being members of that set. As long as the explanation of why
they formulate a single doctrine is missing, the formulation is still arbitrary until the metaphysical paradigm thesis is offered to explain it.

I shall not go into further detail about the minimal thesis or the metaphysical paradigm thesis in this section. What I am attempting to show here is that MN defined by HPA can be a metaphysical doctrine as many of us would expect it to be. The exact content of the two theses, such as what the metaphysical commitments are in the case of the metaphysical paradigm thesis, will be left to be addressed in the next section.

It can be asked whether the minimal thesis or the metaphysical paradigm thesis is the better characterisation. But I do not believe that one of them is more preferable than the other, because I believe that they are equally common and useful understandings of MN. After all, I do not believe that one single true definition of MN has to be singled out. In my view, ‘MN’ is merely a term and it could have multiple meanings; and it is not clear that there is one single correct way to understand it. What is important is simply to sort out which ones are more attractive, defensible, and useful, and to provide defensible regimentations of them. Following the purpose of this dissertation, which is to determine the relation between MN and the categorical ignorance thesis (CI), the relations of both the minimal thesis and the metaphysical paradigm thesis with the categorical ignorance thesis will be discussed in later sections.

2.11. Concluding remarks

I have proposed HPA by using the strategy of SA – which is to search for an extension of current science. I have argued that it can survive the main problems facing DC and the dilemma of the right kind of similarity facing SA, and that it can explain some patterns of behaviour and intuitions of scientists and the folk. And I have suggested two possible ways to develop the proposal into a metaphysical thesis, which are the minimalist thesis and the metaphysical paradigm thesis. Recall that Braddon-Mitchell and Jackson, two proponents of SA, have invited ‘advances in philosophy and indeed in science itself to sort out [the vagueness in SA]’ (2007, pp. 34-35). If HPA is not to be considered as a rival but a development, I hope that it is one advance in response to their invitation, and that it will also provide a framework that allows other advances in philosophy and science to clear things up in an easier way.

Needless to say, there are other attractive approaches to the disciplinary characterisation (DC) in the literature. Because of the limit of space, I cannot address them and compare them with HPA. Those who are interested can refer to Appendix B, in which I will carry out this task.
This proposal of HPA is still incomplete, since HP is a historical concept and there is certainly a problem of historical accuracy. No doubt, there might be historical facts that are in tension with the proposal outlined in this section because the history of science is less linear and unified than it seems. Since this dissertation is not concerned with historical research, it is not and cannot possibly be a detailed study of the actual history of science. The commitments to history are not claimed to be true in terms of a historical study, but rather should be seen as a hypothesis that is waiting to be assessed on the basis of empirical studies in history and philosophy of science. With this in mind, HPA might also have violated the principle of having fewer uncommon or idiosyncratic presuppositions that I outlined in Section 1, since an understanding of the history of science as such is not very likely to be a common belief, even among philosophers. A possible solution I prefer is to see the historical path as a folk concept or even a popular historical myth. According to the solution, what HP is should be assessed in terms of folk intuitions about the history of science, rather than real historical research. This might release the user of the concept of MN from the burden of requiring professional historical knowledge. I am leaving the question of whether the folk concept proposal or the real history proposal is correct as an open question, but for the purpose of this dissertation I will assume the former is correct in order to allow for further assessments of MN. In the next section, I will explore the contents of MN implied by the minimalist thesis and the metaphysical paradigm thesis formulated in this section.
Part I

Section 3. The content of metaphysical naturalism

3.1. Introduction

The purpose of this dissertation is to inquire into the connection between metaphysical naturalism (MN) and the categorical ignorance thesis and determine whether the latter will challenge the former, and hence we have to determine what MN actually is. In Section 2, I proposed an approach to characterising MN, the historical path approach, and I suggested two possible metaphysical readings of it, the minimalist thesis and the metaphysical paradigm thesis. As I explained in that section, my view is that both can be viewed as correct, since they are both common and useful ways to understand MN; and I will thus determine the connections they each have with the categorical ignorance thesis in later sections.

Our inquiry into MN is, however, still incomplete. At the moment, what I have provided with respect to both the minimalist thesis and the metaphysical paradigm thesis are their characterisations, which are only descriptions at an abstract level, but their exact content remains unspecified. To be more precise, these characterisations are basic ideas of what the subject is, not descriptions of the precise details of the subject. One can be fully informed about the characterisations of the two theses but still have no idea whether a thing she is familiar with is metaphysically natural or not with respect to each thesis. Hence, it can be said that they tell us nothing about the actual content of MN. Therefore, at the present moment, the relation between MN and the categorical ignorance thesis cannot be determined.

The purpose of this section is to spell out the content of MN, which includes the content of the minimalist thesis and the metaphysical paradigm thesis.

3.2. The content of the minimalist thesis

According to the minimalist thesis, MN should be seen as a metaphysical thesis that minimises metaphysical commitments. MN characterised as such can be defined as follows:

**The minimalist thesis:** Every entity or property instantiated in the actual world is metaphysically natural – (1) in the sense of (a) being posited by scientific concepts that are
on the historical path or (b) being likely to appear on the historical path in the future, or (2) in the sense of playing only a set of theoretical roles that is closely similar to one played by a metaphysically natural entity or property in the former sense.

Since I have discussed the idea in detail in Section 2, I shall not repeat the explanation of its details again. The very rough idea is that the distinction between the metaphysically natural and non-natural rests on whether they are related to the historical path in some adequate ways mentioned in the definition above; and it is minimalist in the sense that it requires its follower to avoid positing anything else.

Therefore, the content of the minimalist thesis can be said to include two parts: the first consists of the path-related concepts, such as the concept of an electron; the second is the exclusive attitude towards other ontological commitments. I believe that not much can be said about these two parts because of the following reasons. On the one hand, it is practically impossible to list all posits of path-related concepts in this dissertation because of their tremendous number, nor is it necessary, because every reader who is sufficiently informed about science should have an idea of what they are. On the other hand, the exclusive attitude is fairly simple and nothing more can be said about it. With what has been said, the connection between the minimalist thesis and the categorical ignorance thesis can be inquired into in later sections. Hence, the discussion of the content of the minimalist thesis will end at this point.

### 3.3. The content of the metaphysical paradigm thesis

#### 3.3.1. Understanding the metaphysical paradigm thesis

In Section 2.10.2, I introduced the concept of core theories, which are those scientific theories that have most if not all of their main concepts on the historical path. There might be some metaphysical commitments shared by the core theories (or the scientists who work on them). Presumably, these commitments not only are the commonalities between the core theories, but also have explicitly or implicitly served as part of the guiding forces of the path. These metaphysical commitments are considered as a metaphysical paradigm by the metaphysical paradigm thesis. The reader is reminded that the definition of MN according to the metaphysical paradigm thesis is the following:
The metaphysical paradigm thesis: Every entity or property instantiated in the actual world is metaphysically natural – in the sense of being compatible with the metaphysical paradigm shared by the core theories.

Let us call the metaphysical paradigm the core paradigm (CP). The actual commitments of CP are yet to be spelled out, and my aim here is to do so. This is an important step of this dissertation because only with the list completed can the connection between the metaphysical paradigm thesis and the categorical ignorance thesis be determined. It should be noted that the historical path, and therefore the core theories and CP, are historical concepts. Because this dissertation is research into analytic metaphysics, not the history of science, there cannot possibly be a detailed study of the actual history here. For the purpose of this dissertation, what I can do, as I have stated in Section 2.11, is assume another possibility, which is to understand the historical path as a folk concept or popular myth with respect to the history of science. While I will provide my preferred list of commitments of CP, it will not be claimed to be correct in terms of historical accuracy. It is only based on my ‘folkish’ intuitive understanding of the history of science and abstract philosophical reflections, of which the resulting content might in fact be historically inaccurate.

In the following sub-section, Section 3.3.2, I will spell out the commitments of CP. My strategy will be to analyse the major proposed definitions of MN in the literature other than the disciplinary characterisation and see if any of them (or any of the conditions that they consist of) is intuitively (and conceptually) acceptable to be considered as a commitment of CP. The reason for this strategy is that, as I have argued in Section 1.5.2.1, other philosophers’ definitions of MN seem to be the regimentations of what they take to be the content of MN, and hence they are presumptively very useful insights into what the commitments of CP are.

No doubt, my preferred list of commitments of CP is solely based on my intuition with regards to what the history of science is like and abstract philosophical argumentation. In addition, it is presumably not the only possible intuitive way to understand CP. Hence, many readers may disagree. I agree that this is a drawback, and thus I will aim at a weaker and more moderate goal than arguing that it is the correct list. More precisely, the purpose of providing the list is to set out an example of how MN can be understood. And of course, my discussion in the later chapters with respect to its connection to the categorical ignorance thesis should be understood in the same manner.

It is important to note one point before we begin. Many might find it puzzling when I consider a doctrine like ‘there are no divinities’ to be a commitment of CP. This is because many scientists
obviously believe in the opposite. After all, many scientists who work on the core theories are theists, not naturalists. When I consider such doctrines to be commitments of CP, what I mean is that they are taken by the scientists to be their assumptions or working hypotheses when they are working within their discipline, not their personal beliefs. For example, it is a common practice that a theist scientist should set her theistic beliefs aside and assume that there is no God when she is doing her academic job, and this is perfectly consistent with her personal belief that God exists and is the creator of the universe.

One last note. Given the close relation between MN and physicalism, I will also assess proposed conditions of physicalism, even though I do not consider the two doctrines to be identical.

3.3.2. Possible candidates

3.3.2.1. Everything is spatiotemporal

David Armstrong (1978, p. 261; 1997, p. 5) defines naturalism as the following:

Reality consists of nothing but a single all-embracing spatio-temporal system. (1978, p. 261)

Similarly, in Jaegwon Kim’s definition of physicalism, the clause ‘the spacetime world is the whole world’ (2005, p. 150) is included.

No doubt, complaints can easily be made about Armstrong’s view that the space-time he has in mind is ‘a single all-embracing […] system’, because it is inconsistent with the multiverse theory in science, according to which there are initially many universes. I agree that the view is hard to defend. However, what is worthy of discussion here is the view that everything is spatiotemporal, not the single system view.

Nevertheless, the view that everything is spatiotemporal seems to be inconsistent with recent scientific theories as well. While Armstrong believes that space and time can be ‘analyzed’ but not ‘analyzed away’ (1978, p. 263), in some of the latest theories of physics it is simply false that every entity has a spatiotemporal location. In addition, it is unclear that space and time are real. For example, a string in string theory is causal but not spatiotemporal. After all, in my understanding, Armstrong’s doctrine can be said to be a commitment of old-fashioned materialism, which is true of classical physics and was popular until the early 20th century, but it is not up-to-date and is thus not a commitment of CP.
3.3.2.2. *Everything is constituted by microphysical entities*

Many philosophers believe that physics is the study of microphysical entities, and hence for physicalism to be true, everything has to be constituted by microphysical entities (see, for example, Pettit 1993, pp. 213-214; Dowell 2006, p. 39; Kim 2005, p. 150). Michael Tye makes use of this doctrine as one of the conditions of his naturalism. As he puts it:

Higher-level naturalistic phenomena (both token and type) are either ultimately constituted or ultimately realized by microphysical phenomena. (1995, p. 42)

What is the microphysical? Tye does not make this clear, but many others explicitly state that they are (something like) small-sized and atomistic space-time occupiers (see, for example, Pettit 1993, pp. 213-214; Kim 2005, p. 150); and I shall focus on this understanding. The worry of this doctrine, as I have argued in Section 2.3, is that while it might be, much like Armstrong’s spatiotemporalism, an assumption of the old schools of physics, it is no longer assumed the case. For example, it is not obvious that a quantum field is small-sized. Furthermore, according to quantum holism – which many are attracted to – particles are, in fact, not independent entities but instead a single holistic entity (see, for examples of relevant discussions, Teller 1986; Oppy 2000; Schaffer 2003; Esfeld 2004; Ladyman & Ross 2007; Papineau 2008). Therefore, the doctrine cannot be considered as a commitment of CP.

3.3.2.3. *Everything is constituted by (something similar to or represented by) theoretical entities in physics*

Armstrong has an interesting view that naturalism and materialism are significantly different. In his view, while naturalism is a thesis on spatiotemporality, materialism is a completely different thesis:

Contemporary Materialism takes a Realistic view of the theoretical entities of physics – molecule, atom, fundamental particle, and so on – and then asserts that everything there is wholly constituted by such entities, their connections and arrangements. (1978, p. 268)

There are at least two components in Armstrong’s definition, which are (1) realism of physics and (2) the view that everything is constituted by physical entities. I will discuss the latter first and set the former aside at this moment. This is the reason why I added the bracketed phrase ‘(something
similar to or represented by)’ in my version of the doctrine. This is to make sure that the doctrine can be consistent with some sort of instrumentalism and antirealism. More precisely, even though instrumentalists and antirealists do not agree that entities posited by physical theories are real, some of them might agree that something similar to or represented by them is real.

Is this doctrine a commitment of CP? I think yes, because it is quite obvious that core theories in the special sciences, such as chemistry and biology, all have their posits ontologically grounded on those of physics. This, of course, leads to the question of what theoretical entities in physics actually are, since microphysicalism has been rejected in Section 3.3.2.2. There are many proposals in metaphysics and philosophy of science that attempt to provide an idea of what kind of things physics is supposed to study. But I do not take any stance on this matter, nor do I believe that it is appropriate to take any stance here. This is because what is important here is what core theories and scientists working on them assume to be theoretical entities within physics, not how philosophers should characterise physics or theoretical entities within it; and they are, presumably, the list of things that physicists in the physics department are working on. While we all have a rough idea of the list, what has been said is sufficient for our purposes here.

Another problem with the doctrine is with respect to the phrase ‘wholly constituted by’. The worry is that many philosophers consider the meaning of the term ‘constituted’ to be exclusive of emergence, but emergentism is still a widely accepted and respected view in the scientific community. Of course, even though many scientists proclaim to be emergentists, what they have in mind is usually different from what philosophers call emergentism. To be more precise, what they mean is typically the doctrine that some behaviours of a complex system cannot be predicted by using our knowledge of the behaviours of its constituents. This is different from philosophical emergentism, according to which emergent entities or properties ‘arise’ out of their emergence bases and yet are ‘novel’. Nevertheless, the two doctrines are not inconsistent, and it is not obvious that philosophical emergentism is rejected. However, this worry is, after all, merely verbal. If the term ‘constituted’ is understood in a more liberal sense as inclusive of emergence, rather than the way it is typically understood in the philosophical community, then the worry no longer exists.

The last problem is whether my addition of the bracketed phrase ‘(something similar to or represented by)’ into the doctrine is justified. I added it to allow some neutrality towards the debate.

---

14 It is noteworthy that Armstrong might implicitly have microphysicalism in mind, since he only uses particles as his examples. But what is important to our purpose is whether the doctrine he offers can be considered as a commitment of CP, not what he implicitly has in mind.
between scientific realism and some sort of instrumentalism and antirealism; and I believe that the neutrality is required because the actual scientific community does not consist of only scientific realists. In fact, as far as I am concerned, most scientists commit to some intermediate views in between the strongest forms of scientific realism and antirealism. For example, even if a scientist is a realist about physics in general, she might still be an antirealist or an instrumentalist about specific theories, such as some theories of quantum physics, since they are currently fairly difficult to interpret realistically. Yet she can still believe that something similar to or represented by the theoretical entities in quantum physics exist and is the supervenience or emergence bases of the posits of the special sciences. Furthermore, it should be noted that even for the scientists who are realists, most of them are those who accept some weaker versions of realism, according to which our physical theories tell us an approximate truth or a model similar to the truth. This is because, as most scientists agree, our current physical theories are probably incomplete, and our theoretical conceptions of matter can dynamically change as science progresses. Therefore, I believe that the doctrine should be more inclusive than allowing only the strongest version of realism. The bracketed phrase should be included to tolerate weaker forms of scientific realism and antirealism.

3.3.2.4. Scientific realism

Another component of Armstrong’s definition of materialism is realism of physics. Is realism of physics or scientific realism a commitment of CP? I think no. Scientific realism has never been a consensus in science and there is a long history of scientists committing to instrumentalism, according to which scientific theories are not real, but are merely good models that allow us to get predictions right. In Section 3.3.2.3, we also saw that most scientists commit to some intermediate view that is in between the strongest forms of scientific realism and antirealism. In sum, we have no reason to consider the doctrine to be a commitment of CP.

3.3.2.5. Nothing is supernatural

Some philosophers consider antisupernaturalism to be a condition of MN. Graham Oppy has a cluster definition of naturalism, in which he includes the condition that ‘there are none but natural causes involving none but natural entities’ (2014a, p. 39). Michael Tye suggests that ‘on the naturalist view, the world contains nothing supernatural’ (1995, p. 39). Steven Horst also suggests that ‘a naturalistic theory cannot be one that posits the existence of supernatural entities (such as God, angels, or immaterial souls)’ (2007, pp. 15-16).
Unfortunately, none of these authors has made the meaning of the supernatural clear. Tye seems to think that the supernatural is something that is not constituted by the microphysical (1995, pp. 39-40). The problem is, as we have seen in Section 3.3.2.2, that microphysicalism is not attractive because many physical theories seem to be in conflict with it. In Oppy’s other work (2013, p. 6), he seems to be suggesting that the term ‘natural’ in this condition is an opposition to ‘supernatural’, but he has purposefully (and, in my view, reasonably) left the distinction of natural and supernatural unexplained because it is not his main purpose, which is to inquire into topics in philosophy of religion. For Oppy’s purpose, a definition that can distinguish religious (and spiritual) and non-religious (and non-spiritual) worldviews is sufficiently useful.

Despite the fact that the meaning of the supernatural is unclear in the philosophy literature, I consider it to be pretty obvious that ‘nothing is supernatural’ is a commitment of modern science. In fact, in the areas of history and philosophy of science and philosophy of religion, there is a consensus that mainstream modern science (and scientists) has a commitment to a doctrine called methodological naturalism. It is noteworthy that despite its sharing of the same title with a doctrine in philosophical methodology, according to which philosophy should be seen as an extension of science and should therefore use similar methods, the two doctrines have completely different meanings. According to the doctrine, when the scientist attempts to explain a natural phenomenon, she should not appeal to any supernatural entity, especially divinity, even if she personally believes in their existence (see Draper 2005; Fales 2007; Ruse 2010; see also Appendix A.2.5). No doubt, this doctrine is contestable and controversial, and has attracted many criticisms, especially from religious philosophers, but almost all of these criticisms target its justifiability, not its status as the principle or dogma of the mainstream scientific community. The matter of concern is then whether methodological naturalism can therefore be said to be a commitment of CP. I believe that the answer depends on how the supernatural is understood.

The simplest way to understand the supernatural is to understand it as a category opposite to the metaphysically natural. But this is problematic. If MN is defined by CP, saying methodological naturalism is a commitment of CP means that an avoidance of posits violating CP is a commitment of CP. This is circular and trivial. Nevertheless, there are two ways to avoid the circularity.

The first way is as follows. Recall that the metaphysical paradigm thesis is only one way to characterise MN, not the only possible way. With this in mind, when the supernatural is defined as the opposite of the metaphysically natural, the ‘metaphysically natural’ in this context can be characterised in some ways other than the metaphysical paradigm thesis. One example is the
minimalist thesis we have discussed earlier, which defines the metaphysically natural with reference to the historical path without the use of CP. Let us focus on the way a conception of the supernatural (that is relevant to the core theories and scientists who work on them) can be developed by the minimalist thesis. Recall that I argued in Section 3.2 that some scientists have followed some research directions that have to be explained by their attachment to the historical path, and the attachment might be understood as a form of theoretical conservatism. What this seems to show us is that the supernatural might be understood as something in conflict with the scientist’s theoretical conservatism.

The second way to avoid the circularity is as follows. Even if the meaning of the supernatural is yet unclear, we might simply not need to determine the actual meaning of it. What is important for our purpose is what people, especially scientists, take to be members of the category of the supernatural. For example, if most scientists take divinities and the fundamentally mental to be supernatural, then we can simply consider the doctrine that there is nothing divine or fundamentally mental as a possible commitment of CP. This can then prevent the circularity and triviality of directly understanding the supernatural as a category opposite to the metaphysically natural.

In the both ways I discussed, the doctrine ‘nothing is supernatural’ is explained away by other doctrines, which are theoretical conservatism and the avoidance of, say, divinities and should thus be separately assessed. Hence, I will not include antisupernaturalism among the commitments of CP.

3.3.2.6. Theoretical conservatism

Theoretical conservatism is the doctrine that one should make an effort, if not try one’s best, to keep one’s paradigm or theoretical framework in the face of objections and rival views. A good example of the doctrine is a case that we have already come across: in Section 2.8, I argued that cases like the mentalistic interpretation of the Copenhagen interpretation can be considered as sudden divergences from the historical path, and some scientists might be motivated to propose alternatives to such divergences. This motivation, if rational and reasonable, might be a manifestation of theoretical conservatism. What then interests us is whether theoretical conservatism is a commitment of CP.

The doctrine may appear to be an attractive candidate, because it is a common knowledge, even for the folk, that scientists often have a conservative attitude towards new theories, and hence a new theory often requires a certain amount of time to gain wide acceptance and overcome criticism from the mainstream. For example, while Albert Einstein’s theory of relativity did not immediately gain
wide acceptance, Einstein himself also reacted against a probabilistic universe described by quantum mechanics and left his famous phrase, ‘God does not play dice’.

However, I believe that there is, from an intuitive point of view, no evidence that theoretical conservatism is a widely accepted doctrine among scientists associated with the core theories. In fact, most people, scientists or not, do not change their views on and approaches to matters easily, but this does not imply that they all have accepted theoretical conservatism. In most of these cases, their attitudes can actually be explained by psychological and pragmatic factors instead of the doctrine. To be more precise, it might simply be psychologically too difficult or pragmatically too costly for some to change their views and approaches. In addition, the scientist might simply have found insufficient evidence for the new theory, even though we know that it turned out to be the case that the theory was eventually widely accepted. This is because it often takes time for a new theory to gather and demonstrate compelling evidence. I have suggested earlier that some scientists might want to avoid sudden divergences from the historical path, but it is not obvious that the attitude is shared by the majority. Therefore, I find no evidence that theoretical conservatism is a commitment of CP.

3.3.2.7. Nothing is divine, or sacred, or worthy of worship

Another condition of Oppy’s naturalism is that ‘there is nothing in the natural world that is divine, or sacred, or worthy of worship’ (2014a, p. 39). According to Oppy, he sets this condition specifically to exclude pantheism and panentheism because it seems ‘proper’ (2014a, p. 39). It should be noted that Oppy’s definition seems to have a specific purpose, which is to capture the characteristics of a typical atheistic, non-spiritual, and non-mystical worldview for the purpose of studying philosophy of religion; and this purpose is not something we share.

Divinity, sacredness, and worthiness are different things. Let us begin with the first one, which is also the most important one. Even though Oppy’s purpose is not relevant to this dissertation, I believe the exclusion of divinity is, indeed, an important commitment of CP. In my discussion of the doctrine of supernaturalism, I suggested that methodological naturalism (as the term is typically used in science), which is the view that supernatural explanations should be avoided, is a consensus in modern science. In fact, it is pretty obvious in history that the kind of supernatural explanation with which modern science has had the biggest struggle is theological explanations. I believe that the avoidance of theological explanations can be rephrased as an avoidance to posit things that are divine.

Certainly, it can be questioned what the exact definition of divinity is. But our purpose here is to spell out the commitments of CP, and the conventional understanding of the term is sufficient for that.
Hence, the question can be set aside and left for relevant discussions in metaphysics or perhaps sociology and religious studies.

Some may wonder whether this commitment is already implied by the previously discussed commitment that everything is constituted by (something similar to or represented by) theoretical entities in physics, since divinity is supposedly something not constituted by such entities. But I think the answer is no, because we have a very liberal conception of theoretical entities in physics, which only understands such entities as those on the physicist’s list of posits. Hence, divinity might be a theoretical entity in physics, if there is not a commitment against it. Henceforth, there will be some other possible commitments of CP that face the same objection and the same response can be provided, so I shall not repeat the discussion in this paragraph.

One last point. I will exclude Oppy’s mention of sacredness and worthiness of worship from the doctrine because it seems to be a matter of moral value rather than the metaphysical matters that are central to this doctrine (if the doctrine is understood as a commitment of CP). In a later sub-section, Section 3.3.2.12, I will discuss the doctrine that there are no non-constituted moral properties, and I expect that the sacredness and the worthiness of worship Oppy has in mind are examples of this. I will consider the avoidance of positing them as what is implied by that doctrine and set them aside in the following discussion.

**3.3.2.8. There are no fundamental, nonlocal, or ‘early’ mental properties**

Oppy also suggests in his cluster definition of naturalism that ‘the distribution of minds in the universe is late and local: only recently evolved creatures have minds and mental properties, and those minds and mental properties are tied to relatively complex biological structures of the evolved creatures in question’ (2014a, p. 39). Similar doctrines are also commonly seen in proposed definitions of physicalism in the literature, as full definitions or as conditions within them (see, for example, Spurrett and Papineau 1999; Papineau 2005; Montero & Papineau 2005; Wilson 2006).

I believe that an avoidance of positing fundamental or spatiotemporally unlimited mental properties is an obvious metaphysical commitment of modern science, but the avoidance of positing ‘early’ mental properties is not obviously so. Let us begin with the former two. Since Cartesian physics, there has been an assumption that matter is mechanical as opposed to the mental. While the mechanical view of the universe is no longer held by contemporary physics, all core theories inherited the non-mentalistic spirit and share the avoidance of positing mentalities, with the exception of complex functional structures such as our brain. Although some scientists working on the core
theories, such as Newton, believe in panpsychism, idealism, or some other spiritual beliefs, these beliefs always remain personal and have not been able to enter the core theories. Therefore, I consider the avoidance of positing fundamental or spatiotemporally unlimited mental properties as a commitment of CP.

Nonetheless, I do not accept Oppy’s rejection of ‘early’ mental properties as a commitment of CP, even though most scientists take the mind to be a product of evolution. This is because there are no obvious commitments against a Boltzmann brain, an imaginary entity named after 19th century physicist Ludwig Boltzmann. It is a mind that arises out of a physical structure that is arranged randomly in a state of chaos. In my view, science does not rule out an ‘early’ mental property as such, but, instead, has a neutral attitude towards it. For example, if it is discovered that there is a Boltzmann brain that came into existence because of random quantum distributions ten seconds after or two seconds before the big bang, it does not seem to be anything more surprising than the discovery of, say, a new kind of planet in the universe, with the exception that it is more improbable (and is perhaps, for many, a more interesting thing to hear about). The reason for this seems to be that the Boltzmann brain is still local and is constituted by ordinary physical matter that is more fundamental than itself. Meanwhile, other ‘early’ minds, such as God or angels, seem to be against the assumptions of the core theories because their mentalities are fundamental and/or not local. Therefore, I believe that what is expected to be excluded by the doctrine against ‘early’ mental properties is already excluded by the doctrines against fundamental mental properties and nonlocal mental properties. Hence, I will not include the exclusion of ‘early’ mental properties as a commitment of CP.

3.3.2.9. There are no vital forces

David Papineau (2005) argues that being non-vital is a condition of being physical. There are a lot of vitalists in the history of modern science until the early 20th century, including lots of notable figures who are commonly labelled as materialists, such as Denis Diderot. It is, however, now an agreed fact that there are no vital forces. The history of science has proven that vitalism has completely failed and a more mechanistic conception of organisms has gained victory. Hence, the concept of vital force is excluded by the historical path, since the path is only inclusive of concepts that have descendants. With this in mind, all core theories have to be ones that do not assume there are vital forces, even though some of them might be consistent with vitalism and even share some proponents. Therefore, it can be said that ‘there are no vital forces’ is a metaphysical commitment of CP.
3.3.2.10. There are no fundamental teleological properties

This condition is included in Peter van Inwagen’s cluster definition of physicalism and naturalism (2006, p. 81). I think it is an important commitment for modern science to not include fundamental *telos*, such as those in Aristotelian physics. The shift from teleology to experimental science can actually be seen as the beginning of modern science. So ‘there are no fundamental teleological properties’ should be considered a commitment of CP.

It might be objected that *telos* are a kind of mental property, and hence this doctrine overlaps with the doctrine that nothing is fundamentally mental. But I believe that that this objection is not obviously true and is fairly controversial. Hence, I am leaving the two separate, and the controversy will be set aside.

3.3.2.11. All properties are ‘descriptive characteristics’ and ‘factual properties’

In his article ‘Why naturalism?’ (2003), David Copp lists several common definitions of MN in the literature, even though he rejects most of them. Some of those possible definitions are attractive, and thus I will discuss them regardless of Copp’s disagreements with them.

One possible definition of naturalism Copp considers is that all properties are ‘descriptive characteristics’ and ‘factual properties’. He traces this definition back to the literature of Moore, Hare, and Goldman (p. 183). A problem is that it is difficult to see how there can be any ‘non-descriptive characteristic’ and ‘non-factual property’, and it is even more difficult to see what these terms could mean. Hence, this doctrine seems to be trivial. As Copp argues, while a non-naturalist moralist says that someone is a good person, she is also making a description in the ordinary sense. Therefore, I believe that the doctrine is not worthy of any further discussion. However, it seems that in the metaethics circle the terms ‘descriptive characteristics’ and ‘factual properties’ are often taken to mean non-moral properties. A modified version of this doctrine, according to which there are no fundamental moral properties, is worthy of consideration.

3.3.2.12. There are no non-constituted moral properties

Even though it may appear that there are no non-constituted moral properties in the core theories, I believe that the case of this doctrine is somehow unique and controversial. This is because the existence and the nature of moral properties are irrelevant to most areas of natural science. This is probably why the famous evolutionary biologist Stephen Jay Gould – in his seminal and widely read
article ‘Nonoverlapping magisteria’ (2001/1997) – argues that science is the study of facts, whereas religion is the search for ethical values, and that ‘each subject has a legitimate magisterium, or domain of teaching authority – and these magisteria do not overlap’ (741). It is probably also why his view attracted influential followers in the philosophical community, such as Michael Ruse (2010). Setting the matter of religion aside, since it is not our concern here, the worry is that it would be a categorical mistake to say that the core theories assume that there are no such things when the fact is that it is not the purpose of their discipline(s) to be concerned with such things.

I believe that the right answer to Gould’s argument as follows: if his view is that ethical values are never the concern of natural science, then it is simply a mistake. Of course, natural science might not provide us with answers to each moral question, such as whether someone facing the trolley problem should pull the lever, but this does not imply that the basis of moral value cannot be a subject inquired into by natural science. In fact, since Charles Darwin’s seminal book The descent of man (2004/1871), there has always been a metaphysical commitment in the core theories in the areas of evolution of human and animal psychology that morality is an evolved social behaviour of human beings and perhaps some other social animals, not non-constituted moral properties (or entities). Therefore, I consider the doctrine that there are no non-constituted moral properties to be a commitment of CP.

3.3.2.13. All are linked in a causal order

Another possible definition of naturalism that Copp suggests but disagrees with is that the universe of events and states of affairs are linked in a causal order (2003, p. 183). Copp provides an objection to this view, according to which ‘it cannot be assumed that the natural order is causal’ (183) because the fundamental level of reality that realises the causal order might be non-causal in nature. Our concern here is, however, not whether the doctrines we assess are philosophically defensible, but whether they should be considered as commitments of CP.

Nonetheless, in current physics, there is a problem facing the doctrine that is different from Copp’s worry. The worry is, as Daniel Nolan (2015) points out, that while it is almost a universal assumption among metaphysicians that connections between distinct entities, especially behavioural connections, must be causal, it is not the case in the scientific community. For example, quantum gravity is not supposed to be causal. Consider also the EPR experiment in quantum physics. In the experiment, the chance of a measurement turning out a certain way depends on whether we have the information of a former experiment. Many scientists consider the relation between the first and the
second experiments to be non-causal because they can be set in such a way that only a signal faster than the speed of light, which is impossible according to the theory of relativity, can be sent from the first to the second. In fact, since theoretical physics has been increasingly abstract since the 20th century, a lot of physicists do not value the concept of causation but, instead, focus on correlations of measurement qualities and quantities of matter. Therefore, the doctrine that all are linked in a causal order, much like Armstrong and Kim’s spatiotemporalism, might be not up-to-date, and hence I am not including this in my list of commitments of CP.

3.3.2.14. Everything can only be empirically known if it can be known

This is Copp’s preferred definition of the metaphysically natural:

A property is natural if and only if any synthetic proposition about its instantiation that can be known, could only be known empirically. (2003, p. 185)

Copp’s definition is an interesting example of using an epistemic doctrine to define the metaphysically natural. As he writes, ‘my proposal distinguishes natural from non-natural properties on the basis of the nature of our epistemological access to them’ (2003, p. 185). Although his version of naturalism is an ontological thesis, its definition relies on epistemology and methodology rather than ontology.

Copp makes use of the analytic/synthetic distinction and the empirical/a priori distinction in his definition, and hence I should roughly introduce these ideas before we go on. In the standard understanding of contemporary analytic philosophy, analytic propositions usually mean those propositions that can be known simply by knowing what words mean. One example is ‘all bachelors are unmarried.’ On the contrary, synthetic propositions usually mean those propositions that can only be known by knowing the combination of what words mean and some facts about the world. One example is ‘Obama is the current president of the U.S.’. Any given proposition is knowable a priori if it can be known independent of any concrete experience, such as sensual experience and perceptual experience, whereas a proposition that is empirically knowable is knowable through some concrete experiences or some other propositions that are knowable through some concrete experiences.

What Copp’s definition of naturalism suggests is the following. If there is entirely new knowledge that is not derived from older knowledge, then we have to know it through some concrete experiences. Hence, everything can only be known through the basis of concrete experiences. No
doubt, Copp’s naturalism cannot count many presumptively metaphysically non-natural things as
metaphysically non-natural, but this is not a worry for him. This is because his sole purpose is to
deﬁne naturalism in the context of metaethics, and his main opponents in the relevant debates are
moral non-naturalists who believe that there are moral facts that are independent of empirical facts
(e.g. Kantians), and hence the deﬁnition might be a sufﬁciently useful one for his purpose. Contrary
to Copp’s concern, our concern is whether the doctrine is a commitment of CP. I think yes. It is
obvious that modern science has developed to treat empiricism as a methodological norm, and thereby
strongly rejects any synthetic a priori method. In addition, most scientists take their subject matter to
be solely susceptible to investigation.

3.3.2.15. All basic properties are numerically quantiﬁable

Peter van Inwagen has this doctrine as a condition of his deﬁnition of physicalism and naturalism
(2006, p. 81). Unfortunately, it is not very clear what being numerically quantiﬁable means. What I
will do is to assess some possible meanings of it.

First of all, van Inwagen may mean the reducibility of all theoretical descriptions to a purely
mathematical model. Indeed, this kind of view of the universe is shared by many physicists, including
Albert Einstein, who notes that ‘it is [his] conviction that pure mathematical construction enables us
to discover the concepts and the laws connecting them, which gives us the key to the understanding
of nature’ (1934, p. 167). Nevertheless, this view remains controversial among scientists and
philosophers. On the one hand, it is almost a consensus that not all successful theories in special
sciences are reducible to purely mathematical models. On the other hand, even in the case of physics,
it is not obvious that every good physical theory has to be reducible to a purely mathematical model.
The marriage of theoretical physics and mathematical physics is rejected by many physicists. No
doubt, reducibility of all theoretical descriptions to a purely mathematical model is paradigmatically
considered as a virtue, much like simplicity is. Very few scientists would disagree. But standards of
theoretical preference that are a matter of degree and allow exceptions are not the kind of
metaphysical commitment we are searching for.

Perhaps van Inwagen, instead, means being countable by being numerically quantiﬁable. But if
so, then it is so broad that nothing seems to be a possible exception, and hence is almost trivially true.
In sum, it is diﬃcult to see why the doctrine that all basic properties are numerically quantiﬁable is a
commitment of CP.
3.3.2.16. *Everything is a consequence of a network of basic properties and laws*

David Chalmers has this condition in his definition of naturalism (1996, p. 157). First of all, it is noteworthy that there are two possible readings of the sentence ‘everything is a consequence of a network of basic properties and laws’. The first is that the *two groups* ‘properties’ and ‘laws’ *together* constitute the ‘network’ in the sentence. In this case, there must be *both* ‘properties’ and ‘laws’. The second is that the ‘network’ consists of things that can be categorised as members of the *one group* ‘basic properties and laws’. In this case, it is not necessary to have both ‘properties’ and ‘laws’. Regardless of Chalmers’s intended meaning, I will focus on the second possible meaning here and will leave the discussion of the necessity of accordance with laws to Section 3.3.2.17.

I do not consider a literal understanding of the doctrine to be attractive, because it is almost trivially true. The reason is as follows. It seems that all concrete properties can in principle only be categorised as either basic properties or non-basic properties that are composed of basic properties, with the only possible exception being emergent properties; and the same can be said about laws. It is difficult to see how there could be other cases. Even if God exists, it seems that His properties should either be basic, composed, or emergent, since it is difficult to see what would disqualify properties like omnipotence from these categories. Chalmers’s doctrine can exclude emergent properties and laws, but I do not consider anti-emergentism to be a metaphysical commitment of modern science.

Perhaps Chalmers means something posited by fundamental physics by the term ‘basic’. But then the doctrine overlaps with the previously discussed doctrine that everything is constituted by (something similar to) theoretical entities in physics. Therefore, I do not consider this doctrine to be an (independent) commitment of CP.

3.3.2.17. *Everything operates according to laws of nature*

This is another condition of naturalism that Chalmers (1996, p. 157) proposes, and one that is shared by John McDowell (1994) and discussed by J. L. Schellenberg (2007), even though Schellenberg does not think that it is successful. A problem with this doctrine with respect to metaphysics is that it is unclear what laws of nature are and, particularly, whether regularities can be counted as laws of nature. If the answer is no, then the doctrine is simply false. This is because some philosophers of causation believe that there are no laws of nature but only regularities. In fact, this is one of Schellenberg’s objections to this doctrine. If the answer is yes, the doctrine might be too easily
satisfied if not trivially true. For example, suppose that God necessarily punishes evil and that ghosts always avoid light. Can ‘laws’ like these cases be counted as laws of nature? The answer seems very unclear to me. In addition, as far as I am concerned, contemporary philosophical literature on laws of nature does not seem to be useful in helping us to determine the answer of this question. This is because most work has assumed a worldview provided by current natural science and ignored cases like God and ghosts. More precisely, metaphysicians and philosophers of science usually use ordinary cases and current scientific theories as test cases of their theories only.

Nonetheless, the above controversies in metaphysics can be avoided in the consideration of CP. Since the purpose here is to determine whether the doctrine is a commitment of CP, what is important is what the core theories or the scientists who work on them typically assume, not how metaphysicians interpret the doctrine. It seems to me that when scientists talk about laws of nature, they mean some non-teleological behavioural regularities of things, characterised by their somehow repeating patterns of behaviour and manifestations of nature. For example, an electron has a stable quantity of mass and will stably manifest causal powers that are correlated with it in some spatiotemporal locations that are correlated with its previous locations. The exact metaphysical nature of these regularities is not a matter of concern. This simple understanding can avoid the above difficulties in understanding the concept of a law of nature. The question is then whether the doctrine that everything operates according to some non-teleological regularities is a commitment of CP. I believe that it is indeed a common assumption of the core theories and among scientists who work on them.

3.3.2.18. There is a uniformity of everything

Frank Jackson advocates this doctrine as a definition of physicalism. His ostensive definition (i.e. definition that works by pointing out examples) of physicalism is a version of this doctrine, and it is also discussed by Copp as a possible definition of naturalism (Jackson 1998, p. 7; Copp 2003, p. 183). Jackson states:

> 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 and relations, they mean the kinds of properties and relations needed to give a complete account of things like them. (1998, p. 7)
Elsewhere in a co-written work by Jackson and David Braddon-Mitchell, the doctrine is called a ‘uniformity redefinition’ (2007, p. 33) because the core idea of it is that ‘nature is uniform within and without the mind’, as opposed to the dualist view that ‘things in the mind are very different from things in the rest of the world’ (2007, p. 33). To put this simply, every actual thing only has the kind of property that non-sentient objects – such as tables, chairs, and mountains – have. Gregg Rosenberg has a similar view of naturalism, even though he is not a friend of ostensive definition. In his view, naturalism is a requirement ‘to place human beings in the world without making special, ad hoc assumptions that are discontinuous with everything else we have good reason to believe about nature’ (2004, p. 8). He believes that the doctrine is methodological, apparently because he believes that it is an approach to philosophy, but there is still an obvious metaphysical conception within, because it is not just about how we should inquire into the world, but also an assumption about what the world is like.

Understanding the similarity of entities, properties, and relations is difficult, but I believe that the doctrine is, indeed, a commitment of CP. Modern science always assumes that everything in the world is similar in a certain sense. It is generally assumed that any new discoveries will be somehow similar to those previously discovered; and any particular thing or event is assumed to be explicable in similar ways to other particular things or events. Given our current understanding of the physical universe, significantly dissimilar properties and relations like those in the Aristotelian universe are not expected to be discovered; and this does not seem to be merely motivated by the calculation that radically different explanations of cases are much less probable. In fact, the core theories rarely posit two or more fundamental kinds of things.

The only modification needed if Jackson’s doctrine is being considered is that there is no need to specifically point to exemplars of non-sentient objects as examples. This is because our purpose here is different from Jackson’s. Unlike Jackson, I am not attempting to propose a single line of definition of physicalism or naturalism here; instead, I am searching for the commitments of CP. In Jackson’s case, without some physical thing being ostensively pointed to, a uniform universe can be one that is uniform in the sense that everything within only has the kind of property that, say, a ghost has. However, the uniformity doctrine, if understood as a commitment of CP, does not really need to avoid situations like that. All that is needed is the spirit that everything is uniform in some sense, and the way it is uniform can be left to be described by other commitments.

J. L. Dowell raises an objection against the doctrine. In her view, there are cases in physics that some entities and properties are significantly different from the others, such as dark matter and the
property of being it, but it is difficult to see how the discovery of dark matter, for example, is in any sense problematic to anyone (2006, p. 34). Hence, it might be difficult for critics of the doctrine like Dowell to see how it can be plausible.

This objection rightly points out the diversity of things posited by modern physics, but I do not consider it to be compelling. This is because examples like dark matter and the property of being it do not seem to be able to challenge the doctrine. On the contrary, they should still be considered as belonging to the same kind of entities or properties that tables and chairs belong to. This is because they are described by the standard theoretical frameworks in physics – such as Einsteinian theory of relativity, quantum physics, and so on – and have standard theoretical roles such as the possession of mass properties, the possession of gravitational forces, and so forth.

Taking this into account, I believe that real irregularities (in the sense required by the doctrine) occur only when there are entities or properties that require a fundamentally different theoretical framework to describe, even when their constituents are taken into account. Examples of these would include vital forces and ghosts, which act in ways that cannot be described and predicted by the physical laws that are attributed to other things and, in addition, cannot be described by laws similar to the physical laws that are attributed to other things.

One last point. Can this doctrine explain the previously discussed commitments of CP that there are no divinities, fundamental and nonlocal mental properties, and vital forces? I believe that it can, since those things obviously does not have a uniformity with things like chairs and tables. The problem is then whether we still need those commitments when we have this one. I believe that the answer is yes. The reason is as follows. From an intuitive understanding of science, it seems that those commitments are independently held. If physicists were to discover that there are many kinds of things in the world that have no uniformity at all, it seems that science will still resist positing things like divinities and the fundamentally mental. Hence, it seems that it is more preferable to see the commitments as independent.

3.3.2.19. There is a uniformity of the explanatory apparatus of everything

John Searle has an understanding of MN that is similar to Jackson’s ostensive definition. When he explains his ‘biological naturalism’, according to which consciousness is an irreducible biological feature but is nevertheless natural, he explains the concept of naturalism as follows:
[C]onsciousness is part of the natural world along with other biological phenomena such as photosynthesis, digestion or mitosis, and the explanatory apparatus we need to explain it we need anyway to explain other parts of nature. (2007, p. 329)

I shall not discuss Searle’s view(s) in philosophy of mind and shall focus on his definition of naturalism. It is obvious that the clause ‘along with other biological phenomena such as photosynthesis, digestion or mitosis’ is, like Jackson’s ostensive physicalism, a version of the ostensive definition and the uniformity thesis. But it is noteworthy that there is another clause, ‘the explanatory apparatus we need to explain it we need anyway to explain other parts of nature’. Hence, in Searle’s view, there is not only a uniformity of ontology, but also a uniformity of methodology.

I do not consider the doctrine to be a commitment of CP because I find it difficult to understand what Searle means by ‘the explanatory apparatus to explain other parts of nature’. More precisely, I find all possible ways to understand the statement pretty implausible. Firstly, it might mean constitution. If this is the case, the doctrine would be something like the previously discussed doctrine that everything is constituted by (something similar to) theoretical entities in physics, since most agree that biological structures are constituted by entities described in physics. In which case, the two doctrines overlap and there is no need to consider Searle’s doctrine as independent. But Searle explicitly denies this. He points out that the physical is causally explainable by microphysical processes, while the mental is not explainable by physical processes (2007, p. 331).

Perhaps Searle means our methods of inquiring into the causal structure of the entity of concern, which include its features – such as its causal dispositions, the arrangements and causal dispositions of its constituents, and so on. But the problem is then what kind of method of inquiry he has in mind. If it is a very precise methodology, such as ways of making measurements, then it is unclear how a common methodology can be shared between the study of the physical and the study of the mental. This is because of the following two reasons.

Firstly, this seems to be incompatible with Searle’s own beliefs, since he believes in the existence of qualia and the ontological subjectiveness of consciousness (2007, p. 326). To be more precise, if consciousness is irreducibly subjective, then it is unclear how precise methodologies like those used to measure it (if it can be measured) can be shared with other things. Secondly, it is pretty unclear that a common methodology is shared among different areas of science. For example, it is pretty unclear that ecology and microphysics share the same method of investigation. If this reading of Searle is true, the objects that ecology and fundamental physics study, according to the doctrine,
cannot both fall into the category of natural objects. No doubt, the objects which ecology studies can be counted otherwise, because their constituents can be studied by fundamental physics in the way the objects that fundamental physics studies can. But this then collapses into the first possible reading of Searle, which we have discussed and repudiated.

Perhaps ‘apparatus’ does not mean a very precise methodology but the basic epistemic virtues in doing natural science, such as simplicity, productivity, explanatory powers, falsifiability, and so on. However, the doctrine would be empty if this is the case, because these virtues are too general. More precisely, they seem to be consistent with and irrelevant to any kind of metaphysics and ontology and hence cannot be exclusive of any of them.

Given the above difficulties in understanding Searle’s doctrine, I do not consider it as a commitment of CP.

3.3.3. A list formulated

Recall that our purpose here is to spell out the metaphysical commitments of the core paradigm (CP), which is the content of MN as the doctrine is defined by the metaphysical paradigm thesis. The possible commitments of CP I have come across are the following:

1. Everything is spatiotemporal.
2. Everything is constituted by microphysical entities.
3. Everything is constituted by (something similar to) theoretical entities in physics.
4. Scientific realism.
5. Nothing is supernatural.
6. Conservativism.
7. Nothing is divine, or sacred, or worthy of worship.
8. There are no fundamental, nonlocal, and ‘early’ mental properties.
9. There are no vital forces.
10. There are no fundamental teleological properties.
11. All properties are ‘descriptive characteristics’ and ‘factual properties’.
12. There are no fundamental moral properties.
13. All are linked in a causal order.
Everything can only be empirically known if it can be known.

All basic properties are numerically quantifiable.

Everything is a consequence of a network of basic properties and laws.

Everything operates according to laws of nature.

There is a uniformity of everything.

There is a uniformity of the explanatory apparatus of everything.

Among the above possible commitments, I consider only the following ones as the actual commitments of CP: (Note that some have been modified.)

1. Everything is constituted by (something similar to) theoretical entities in physics.
2. Nothing is divine.
3. There are no fundamental and nonlocal mental properties.
4. There are no vital forces.
5. There are no fundamental teleological properties.
6. Everything can only be empirically known if it can be known.
7. Everything operates according to laws of nature.
8. There is a uniformity of everything.

The reader is also reminded that the above list is solely based on my intuition with regard to what the history of science is like (with the addition of abstract philosophical argumentation). Hence, it is provided only as an example of how CP can be understood, but is not to be supposed as a historically accurate claim of what CP is like, nor is it the only possible intuitive way to understand it.

3.3.4. Some additional remarks on the metaphysical paradigm thesis

Some might object that the list is unreasonably large and has thus violated the principle of simplicity. I agree that the list is larger than almost all definitions of MN in the literature, but I do not consider this to be a real failing because there is a good reason for the size of the list. After all, the metaphysical paradigm thesis is not as complex as it seems, since being consistent with CP is the only condition of being metaphysically natural. The commitments in the list are merely its actual details. In addition, a large collection of assumptions is what a paradigm is supposed to be, and hence the long list of the commitments of CP is simply natural. Recall that some authors, such as Chalmers and Oppy, defend
cluster definitions of MN (see Section 2.3). In their view, a thing is metaphysically natural because it meets all of several conditions, and, indeed, all of these authors’ lists of conditions are shorter than this one. But as I have argued in Section 2.3, their approach seems to be merely adding conditions arbitrarily and subjectively with the sole purpose of excluding types of things that they do not approve of. The list of commitments of CP, on the contrary, is not arbitrary.

Another possible objection is that the use of the concepts of the historical path, the core theories and the core paradigm in the definition now seem to be unnecessary, since the resulting list of metaphysical commitments seems to be the same as the list of metaphysical commitments held by contemporary science. The worry is that the metaphysical paradigm thesis can simply define MN with reference to the list of metaphysical commitments of contemporary science and still remain the same, while being a lot simpler. I disagree with this objection because I do not think that the list of metaphysical commitments of CP and that of contemporary science are the same. It is important to remember that there is the problem of naturalistically unacceptable scientific theories (see Section 2.4.5). Given the diversity in current scientific theories, if there were no reference to historical theories but merely a reference to current theories, then there would be difficulties in deriving our current list. For example, it would be unclear why fundamental mental properties and fundamental teleological properties were excluded, since there are cases like the mentalistic interpretation of the Copenhagen interpretation. On the contrary, as I have argued in Section 2.8, the historical path approach can point out that naturalistically unacceptable scientific theories, such as the mentalistic interpretation of the Copenhagen interpretation, are sudden divergences from the path.

Another worry of defining MN with reference to the set of metaphysical commitments of contemporary scientific theories is that we might have to include some metaphysical commitments that most of us would take to be irrelevant to the concept of MN, such as eternalism in philosophy of time (which most current theoretical physicists commit to). Without the concept of the historical path, it is difficult to see how we can distinguish relevant metaphysical commitments from others.

Therefore, if we define MN with reference to the set of metaphysical commitments of contemporary scientific theories instead of that of CP, the current list of commitments should, in fact, be replaced by another one, not remain the same; in addition, the replaced list is not what we expect of MN. In sum, there is no reason for such a replacement to take place.

4.5 Final remarks on Part I

The discussions in Part I come to an end. I can summarise what I have done in Part I as follows:
In Section 1, I proposed a method for characterising MN. In particular, the kind of MN that is required by the purpose of this dissertation, which is to determine whether the metaphysical framework shared by many if not most analytic philosophers is correct. I argued that the best method is to see MN as a collection of theoretical roles spelled out in some tacit consensus in standard philosophical discourses and to provide the best regimentation of it.

In Section 2, I proposed a new approach to characterising MN, which is the historical path approach (HPA). I also argued that the approach can have two metaphysical understandings, the minimalist thesis and the metaphysical paradigm thesis. I consider both to be correct.

In this section, I spelled out the content of the minimalist thesis and the metaphysical paradigm thesis. The metaphysical paradigm thesis is considered with weight, since the commitments of the core paradigm (CP) have to be listed.

With what has been done, the concept and content of MN are clear enough. Recall that the purpose of this dissertation is to inquire into the connection between MN and the categorical ignorance thesis and determine whether the latter will challenge the former; our understanding of MN is now ready for this purpose.
Part II

Categorical ignorance
Part II

Section 4. Introduction to categorical properties

4.1. Introduction

The purpose of this dissertation is to inquire into the connection between metaphysical naturalism (MN) and the categorical ignorance thesis (CI) and determine whether the latter will challenge the former. MN is a metaphysical doctrine or framework held by many contemporary philosophers, according to which all actual entities, properties, or relations are metaphysically natural. But the worry is that CI, a philosophical thesis that has recently gained attention through the work of Rae Langton (1998), David Lewis (2009), Frank Jackson (1998), Simon Blackburn (1990), and Michael Smith and Daniel Stoljar (1998),15 might defeat MN by being incompatible with it or by reducing its likelihood.

In Part I of this dissertation, I formulated two characterisations of MN by the historical path approach (HPA). These two characterisations are both candidates of the best regimentations of MN that I consider to be attractive. In Part II, I will introduce and assess CI. I will introduce the thesis, determine whether or not it is true, and consider which arguments for CI are best.

CI can be roughly stated as the following:

**CI:** Categorical properties are in principle unknowable. All of our knowledge is, in fact, about other properties, such as causal and dispositional properties.

Before I explain what categorical properties are taken to be and consider the major arguments for CI, I should note that among the philosophers who discuss CI, only Blackburn (1990) and Smith and Stoljar (1998) explicitly discuss categorical properties. Other major contributors discuss the ignorance of intrinsic properties or fundamental properties, not categorical properties. For the sake of avoiding confusion, I will set CI aside for a moment and, following Langton’s ‘Kantian humility’ thesis and Lewis’s ‘Ramseyan humility’ thesis, call this group of ignorance theses the ‘epistemic humility’ thesis (EH). As this section goes on, I will show that all major versions of EH in the literature

---

15 For an earlier seminal work, see Armstrong 1961.
are actually CI expressed in different ways because the properties they concern are in fact the same; they are actually categorical properties under different descriptions.

EH can be roughly stated as follows:

**EH:** Categorical properties, and/or intrinsic properties, and/or fundamental properties, are in principle unknowable. All of our knowledge is, in fact, about other properties, such as causal and dispositional properties.

We have a set of kinds of properties: categorical, intrinsic, and fundamental. Call the properties that are of at least one of these kinds ‘CIF properties’. What we have is not a single kind of property because the kind of property of concern in EH varies between different authors. All versions of EH have a hidden presupposition, according to which CIF properties exist in the actual world. Assuming the contrary (that CIF properties do not exist), if we have perfect knowledge of all actual properties other than CIF properties, then we know all facts about the world. Hence, the world has no further facts that we are, as the thesis suggests, irremediably ignorant of. EH would then be false, or meaningless because what it can tell us, at most, is that if we (counterfactually) live in a possible (or imaginary) world with CIF properties, we cannot know about them, but it cannot tell us any fact about the actual world.

This section is a review of the literature, with the aim of narrowing down the scope of discussion of Part II. The aims of this section are (1) to roughly explain what the different kinds of CIF properties are, and (2) to explain why all CIF properties discussed in the literature on EH are actually the same kind of property, which is what most contemporary philosophers call ‘categorical properties’. Making these conceptions clear allows us to discuss CI and its relevant literature in the following sections without confusion.

### 4.2. Introduction to CIF properties and the use of them by EH theorists

As we have seen, there are three kinds of CIF properties: categorical properties, intrinsic properties, and fundamental properties. Very roughly, the three kinds of properties as follows:

**Categorical properties:** Properties that are not causal properties or dispositional properties, i.e. causal powers, (e.g. the length of a ruler).\(^{16}\)

---

\(^{16}\) Of course, the length of a ruler does make a causal difference to the world. However, for many, it could be understood
**Intrinsic properties:** Properties that objects have of themselves, independently of their relations with other things (e.g. my having a brain).

**Fundamental properties:** Properties that are perfectly basic in the sense of not being constituted, disjunctive, determinable, negative, and so on. Unfortunately, in a naturalistic worldview, no example can be given because there is no evidence that current physics has reached the fundamental level. However, examples can be taken from non-naturalistic worldviews. According to Berkeleyan idealism, for example, being God’s mind might be a fundamental property.

The above characterisations are all very imprecise and rough descriptions. Like physicalism and metaphysical naturalism, how each of them should be defined in formal terms is being intensely debated in the literature (see Weatherson & Marshall 2012 for a detailed introduction to some ways of drawing the intrinsic/extrinsic distinction that can be found in the literature). But for the purpose of this dissertation, the above rough characterisations will suffice.

Despite the fact that theorists of EH use different terms to describe the properties that are considered in principle unknowable, it seems that (1) they have the same kind of property in mind. On the other hand, (2) all of these characterisations only partially capture what is supposed to be in principle unknowable. My view is that all major versions of EH in the literature are actually concerning the same kind of property. I will also argue that if I am right, all of these versions are actually CI in disguise. In Section 4.3, I will discuss some examples of the discussions of EH in the literature; the reasons for my position should thus become clear.

### 4.3. Examples of the use of CIF by EH theorists

**4.3.1. Lewis on fundamental properties**

When Lewis discusses the ignorance of fundamental properties in his article on EH, ‘Ramseyan humility’ (2009), I believe that he had already assumed his own metaphysical framework, according to which fundamental properties are intrinsic and non-dispositional. Of course, not all philosophers understand fundamental properties in this way. For some philosophers such as James Ladyman and Don Ross (2007), the relational structure of the world is fundamental; and for some philosophers such as a structural property that partially determines the causal properties or dispositional properties of the ruler, such as the way the ruler causally behaves under gravity, but not those causal properties or dispositional properties themselves.
as Sydney Shoemaker (1980) and Chris Swoyer (1982), all fundamental properties are dispositional. However, in his article, Lewis explicitly states that what he calls fundamental properties are intrinsic and not structural (p. 204); and he also distinguishes his view from those of Shoemaker and Swoyer (pp. 220-221n13). In fact, he has explicitly argued against the views of Shoemaker and Swoyer elsewhere (1986a, p. 163).

While Lewis takes the above assumptions to be true, he seems to take them to be contingently (but not necessarily) true. He writes in an earlier work:

Perhaps there might be extra, irreducible external relations, besides the spatiotemporal ones […] It is not, alas, unintelligible that there might be suchlike rubbish. Some worlds have it. […] But if there is suchlike rubbish, say I, then there would have to be extra natural properties or relations that are altogether alien to this world. (1986b, p. x, my emphasis.)

We can see that he explicitly considers the kind of view of Ladyman and Ross to be contingently false, and this probably applies to that of Shoemaker and Swoyer as well if dispositions are considered as a kind of irreducible external relation. In other words, Lewis is only taking actual fundamental properties into account in his version of EH.

4.3.2. Langton on intrinsic properties

Another influential work on EH is Langton’s book Kantian humility: our ignorance of things in themselves (1998), in which she discusses intrinsic properties. Before I introduce her views, I will note that her book is on the history of philosophy. More precisely, it is on how we should interpret the work of the 18th century German philosopher Immanuel Kant. Nevertheless, the views in the book are of independent and contemporary philosophical interest because they are defended and translated into the language of analytic philosophy by Langton. As she explicitly states, discussions of the views are found in the contemporary debate between categoricalism and dispositionalism (1998, pp. 118, 120n32; 2006, pp. 176-177), which is contributed to by contemporary metaphysicians including Frank Jackson, David Armstrong, Sydney Shoemaker, Brian Ellis, C. B. Martin and many others (1998, p. 120n32). Moreover, the book has attracted attention in contemporary metaphysics and epistemology. For example, Lewis writes, ‘my interest is not in whether the thesis of Humility, as she conceives it, is Kantian, but rather in whether it is true’ (2009, p. 203). For the rest of this
dissertation, I will treat Langton’s own view and her interpretation of Kant as if they are the same thing, and when I mention the view of Langton I mean the collection of the two of them.

To avoid confusion, I will note that Langton considers Langton’s Kant’s view to be independent of and inconsistent with Kant’s most influential metaphysical thesis, transcendental idealism (2004, p. 129; 1998, p. 143n7), which is (very roughly) the thesis that the mind-independent reality is unknowable, and that all laws of nature including metaphysical laws, physical laws, and special science laws are, in fact, cognitive laws that conscious and rational agents necessarily possess. Transcendental idealism is irrelevant to this dissertation.

Langton’s Kantian humility thesis is an ignorance thesis about intrinsic properties, but she notes that it does not apply to all kinds of intrinsic properties (1998, p. 62). For example, most consider the roundness of a cup intrinsic, and some also consider negative properties such as ‘not being Mickey Mouse’ and ‘not being number 2’ intrinsic. Langton, in a separate article with Lewis (1998), considers all of the above cases to be intrinsic properties. These properties are apparently knowable because we appear to know about them by scientific means; in addition, Langton makes it clear that she is a realist about science (1998, p. 143n7). Therefore, Langton, in fact, does not consider all intrinsic properties to be unknowable, even though she simply speaks of the ignorance of ‘intrinsic properties’ most of the time. In her view, some ‘comparatively intrinsic properties’ are knowable; only ‘absolutely intrinsic properties’ are unknowable (1998, p. 62). The term ‘comparatively intrinsic properties’ means those that are constituted by extrinsic relations, whereas the term ‘absolutely intrinsic properties’ means those that are not. With this in mind, we can see that the intrinsic properties that her version of EH concerns are fundamental.

Langton also argues that fundamental properties cannot be causal or dispositional, and takes this as an important presupposition of her Kantian humility thesis. I will discuss her arguments in detail in Sections 6 and 7. For now, we can safely conclude that the kind of intrinsic property that Langton’s version of EH concerns is fundamental and categorical.

4.3.3. Blackburn’s use of categorical properties

Blackburn’s article ‘Filling in space’ (1990) is a seminal article that argues for an ignorance thesis about categorical properties. In the article, Blackburn contributes to the debate between categoricalism and dispositionalism. To put the debate very roughly, categoricalists believe that causal properties and dispositional properties require categorical properties to act as their bases, while dispositionalists believe that only dispositional (or causal) properties exist. I will defer more detailed
discussion of this debate until Section 6. The main point here is that Blackburn is a categoricalist, and in his article, he argues that categorical properties are something of which we are necessarily ignorant.

What the terms ‘categorical property’, ‘categorical basis’, and ‘categorical ground’ mean is often not adequately spelled out. Recall that categorical properties are neither dispositional nor causal. For example, the length of a ruler is categorical, and colour properties, such as redness, are sometimes considered categorical.¹⁷ A categorical basis or a categorical ground is a categorical property that grounds a dispositional property. This way to make the distinction is uncontroversial in the literature. However, when philosophers speak of a ‘categorical basis’, it can mean different things. First, this phrase could mean the structural properties that realise the dispositions we find in most cases in science. For example, the information output of a computer is realised by its complex mechanical structure, and the dissolution of salt is realised by its molecular structure. Second, this phrase could pick out non-structural, perfectly intrinsic, and fundamental properties. These properties ground everything else that exists. Examples of them cannot be provided because current science tells us nothing about that sort of property.

The crucial question is whether Blackburn is referring to the ordinary structural categorical properties or some perfectly intrinsic and fundamental categorical properties. He seems to be referring to the latter. He writes, ‘Science finds only dispositional properties, all the way down’ (1990, p. 63). This claim would not be reasonable if he has structural categorical properties in mind, because science obviously can discover the depth of an ocean, the molecular structure of salt, the muscular structure of a human arm, and so forth. Hence, if Blackburn’s view is reasonable, he cannot mean structural properties when he uses the term ‘categorical properties’. In fact, he explicitly draws a connection between categorical properties and the Lewisian notion of fundamental properties:

In Lewis’s exploration of these issues, categoricity comes in the ‘Humean mosaic’ or pattern of perfectly natural intrinsic properties which need nothing bigger than a point at which to be instantiated. (p. 65)

Therefore, we can safely conclude that Blackburn is concerned with fundamental intrinsic properties when he discusses categorical properties.

¹⁷ This is, of course, controversial because of the debates surrounding it in philosophy of mind. I shall not get further into the debate.
4.4. The relation between CI, EH, and CIF

We have seen that the kinds of CIF properties all major EH theorists are concerned with is a single kind of property, which are properties that are categorical, intrinsic, and fundamental. Therefore, it can be concluded that different versions and discussions of EH are actually referring to the same thing. No doubt, it can still be questioned whether theories of and arguments for EH apply only to the kind of property. This is because it is one thing to say that all major EH theorists focus on the kind of property, and another thing to say that their theories and arguments apply only to them.

In fact, even for the major EH theorists, not all of them believe that theories of and arguments for EH apply only to the kind of property, or perhaps even CIF properties. Lewis, for example, has a view which he calls ‘spreading humility’ (2009, p. 214). In this view, most properties about which we should be humble are not fundamental properties, even though fundamental properties are the kind of CIF properties that he discusses. The reason is that some structural properties supervene on fundamental properties, and since we cannot know of their supervenience bases, we cannot have perfect knowledge of them as well. Lewis’s view is a perfect example of a version of EH that, if correct, not only applies to properties that are categorical, intrinsic, and fundamental, but also implies our irremediable ignorance of some other things. Of course, the ‘spreading humility’ view requires a point of departure before the ‘spreading’ is discussed, and thus Lewis’s discussion still focuses on properties that are categorical, intrinsic, and fundamental, but it still shows us that theories of and arguments for EH may apply to other things.

Nonetheless, I will limit the discussion in the following sections to properties that are categorical, intrinsic, and fundamental. This is because it is impossible to discuss all kinds of properties we might be irremediably ignorant of in this dissertation, especially when this dissertation is about not only CI or EH, but also metaphysical naturalism. Taking the limit of scope into account, we need to focus on a particular kind of property. We have seen that properties that are categorical, intrinsic, and fundamental are always the focus of related debates in the literature. I believe that it makes our discussion easier to begin with if we use such standard context as a point of departure.

In the following sections, I will use the term CI to stand for all versions of EH, and I will simply use the term ‘categorical properties’ (or the term ‘categorical bases’ when they act as the bases of dispositional properties) to stand for properties that are categorical, intrinsic, and fundamental (unless specified) because this follows the more standard terminology. However, the reader should be aware that there are, in fact, kinds of categorical properties other than those I discuss.
The following sections of Part II will be as follows. In Section 6, I will argue for the existence of categorical properties. In Section 7, I will further inquire into their nature. In Section 8, I will determine whether CI is true.
Part II

Section 5. The existence of categorical properties

5.1. Introduction

As I have noted in Section 4, not everyone believes that categorical properties (that are not constituted by dispositional properties) exist. Many philosophers believe that there are only dispositional properties that have no categorical bases. If this is true, the categorical ignorance thesis (CI) is false (or meaningless) since it will not really point out anything in the (actual) world that we are irremediably ignorant of. Therefore, to make sense of the discussion of CI, it must be first shown that there are categorical properties. In this section, I will argue for their existence.

In Section 5.2, I will introduce the three major doctrines concerning the existence/non-existence and the nature of categorical properties. In Section 5.3, I will give a rough idea of the kinds of arguments for the existence/non-existence of categorical properties that I will inquire into in this section. The rest of the structure of this section will be left to be introduced in the subsection after we have a better idea of what the sides of the debate and what the arguments we will inquire into are.

For the reasons specified in the previous section, Section 4, it is noteworthy that only those categorical properties that are fundamental and intrinsic, not other kinds of categorical properties, are the interest of this section and what I typically mean by the term ‘categorical properties’ unless specified otherwise.

5.2. Major doctrines concerning categorical properties

In contemporary metaphysics, there are three major doctrines concerning the existence/non-existence and the nature of categorical properties. They are:

**Strong categoricalism:** All fundamental properties are categorical properties without dispositional essences intrinsic to them (see below). For there to be dispositional properties, categorical properties are required to act as their bases. But the relation between the two kinds of properties (i.e., which categorical property is playing which dispositional role) is

---

18 Of course, it will point out that, if we counterfactually live in a possible or imaginary world with categorical properties, we cannot know of them.
determined by the laws of nature – or what described by the laws if the laws are understood to be descriptions – that are extrinsic to categorical properties, not the intrinsic essences of categorical properties (see, for example, Armstrong 1997; Langton 1998; Lewis 1986b, 2009).

**Strong dispositionalism:** All fundamental properties are dispositions, which are bare and thus have no categorical properties that bear them. Laws of nature in this case are the descriptions or analyses of the properties (see, for example, Shoemaker 1980, 1998; Swoyer 1982; Hawthorne 2001; Bird 2005).

**Weak categoricalism/dispositionalism:** Dispositional properties are roles played by categorical properties, or are identical to the categorical properties that bears them. In the former case, the nature of a dispositional role of a categorical property is determined by the intrinsic dispositional essences of the property. The laws of nature, in the former case, are (the descriptions or analyses of) the essences, and in the latter case, are the descriptions or analyses of the properties (see, for example, Mumford 1998; Martin & Heil 1999; Heil 2004a, 2004b, 2005; Strawson 2008b; Oderberg 2009; Schroer 2010a).

By a dispositional essence, I mean any feature or nature of a property that determines its dispositional role. For extrinsic laws of nature, numerous conceptions are on offer. Examples include the governing conception – according to which laws are (descriptions of) something ontologically independent and primitive that ‘governs’ the categorical properties (see, for example, Carroll 1994), the theistic version of the governing conception – according to which the governing laws are set up by and/or supervene on God (see, for example, Langton 1998), and the Humean conception – according to which laws are merely (descriptions of) systematic distributions of categorical properties (see, for example, Lewis 1986b).

The categorisation above is a very rough and not exhaustive one (following a similar categorisation provided by David Armstrong [1997, ch. 5]). The view of each author is somehow different, and thus not everyone’s view can fit into the three categories perfectly. For example, Brian Ellis and Caroline Lierse (1994) have a view on fundamental dispositional properties that is exactly like strong dispositionalism, but they also believe that both fundamental dispositional properties and spatiotemporal properties are indispensable from the constitution of non-fundamental dispositional properties. Spatiotemporal properties, in their view, are irreducible to dispositional properties and are thus fundamental and categorical (p. 39). The uniqueness and complexity of the view makes them
difficult to categorise into the three categories of views I listed. Furthermore, the debate between categoricalism and dispositionalism often involves views regarding the metaphysics of laws of nature, which is itself an independent and complicated philosophical topic being intensively researched. This further increases the diversity of views in the literature. A more sophisticated categorisation may arrange the views into more than three categories (see, for example, Dumsday 2015, pp. 108-109). However, the rough categorisation above is sufficient for my purposes.

5.3. Categoricalism (strong or weak) versus strong dispositionalism

If strong dispositionalism is true, dispositional properties have no categorical bases. My view is that either strong or weak categoricalism is true because I believe that, as both doctrines suppose, dispositions require categorical bases as their bearers. In the rest of this section, I will argue for categoricalism (strong or weak) and defend it against major objections offered by strong dispositionalists.

The following discussion will concern the role and the explanatory value of categorical properties in science. This is because most major arguments against categoricalism belong to a family of argument that I call the \textit{razor arguments}. They argue that only dispositional properties have a role in science or in (philosophical) explanations, whereas categorical properties have no such roles and therefore should not be posited in our ontology. The goal of this section is to show that the assumption of the razor arguments is false. Not only do dispositional properties have explanatory roles in science; categorical properties also have significant and indispensable roles. The explanatory roles I attribute to categorical properties are all what I call \textit{binding roles}, which are roles posited to explain the common features of properties. I will argue that categorical properties have two kinds of binding roles in science: (1) holding the spatiotemporal locations of dispositions together or (2) unifying some features of dispositions. More details will be given as the discussion goes on. I will argue that consideration of the binding roles not only motivates categoricalism and resists the razor arguments, but also shows that strong dispositionalism is untenable because strong dispositionalism cannot provide alternative explanations of the roles that are tenable.

The structure of the rest of Section 5 will be as follows. In Section 5.4, I will outline the razor arguments against categoricalism offered by strong dispositionalists. In Section 5.5, I will argue that categorical properties have two binding roles in scientific explanations. I will also offer a possible explanation of why many theorists are prone to miss them. In Section 5.6, I will critically assess two alternative hypotheses with regard to the binding roles that are available to the strong dispositionalist.
In Section 5.6, I will provide some further considerations of empirical scientific theories, in which I will focus on James Ladyman and Don Ross’s ontic structural realism (OSR).

Note that the debate over which major doctrine concerning categorical properties is true is an important focus of research in the literature; and the debate, as I have noted, is also strongly related to the debate about the metaphysics of laws of nature, which is itself an independent topic that is being intensively researched. It is thus impossible to assess all major arguments for and against each doctrine in a single section. Therefore, rather than proving that categoricalism is true or providing a comprehensive assessment of the three major doctrines regarding the nature of categorical properties and dispositional properties, I am merely attempting to provide motivations for considering categoricalism (strong or weak) to be more tenable than strong dispositionalism and using that conclusion as an assumption for the rest of this dissertation. While I take it to be possible that my consideration of tenability in this section might be overridden by independent arguments, I will bracket them off and assume that they have no force.

5.4. The razor arguments

Most major arguments against categoricalism are what I call the razor arguments. These arguments do not target the possibility of the existence of categorical properties. Instead, they appeal to some epistemic principles or to some considerations of ontological economy and argue that we should avoid positing categorical properties. These arguments can be categorised into two main kinds, which I call (1) the argument from ignorance and (2) the argument from science. I will briefly outline the two arguments and explain why I believe that these arguments are unconvincing.

5.4.1. The argument from ignorance

The argument from ignorance can be found in many seminal works in the area, such as Shoemaker 1980, pp. 116-117, Swower 1982, pp. 204-205, Black 2000, pp. 92-95, and Bird 2005, p. 453. The argument, as I understand it, is based on two (explicit or implicit) epistemological principles. The first is what I call epistemological dispositionalism:

**Epistemological dispositionalism:** All knowledge (we have about concrete entities) is about dispositions and their manifestation (i.e. causal behaviours).

The second is an epistemic principle that I call *the principle against the non-contributive:*
The principle against the non-contributive: We should not posit anything that cannot contribute to our knowledge.

With these two epistemological views combined, since categorical properties are not dispositions or their manifestation, we cannot have knowledge of them, and thus we should not posit them. An example of the argument is provided by Alexander Bird. As he puts it:

We do not want our metaphysics of properties to condemn us to necessary ignorance of them. And so we should reject [...] categoricalism. (2005, p. 453)

What is the motivation for epistemological dispositionalism? For many proponents of the view, we in principle can only know things through causal connections between our sensory states and the things of concern. For example, Sydney Shoemaker writes, ‘Observing something is being causally influenced by it in certain ways’ (1980, p. 116). Hence, every concrete thing we know about has to be causally efficacious. While dispositional properties are causally efficacious, categorical properties are (for many) causally inert. Therefore, we can know about dispositional properties but not categorical properties. This argument will not be assessed in detail at the moment, because similar arguments are offered in the works on the categorical ignorance thesis (CI) by categoricalists such as Langton, Lewis, Jackson, Blackburn, Stoljar, and Smith. Since CI is a main topic of this dissertation, arguments for it will be left for a very detailed assessment in an independent section, Section 8. The above rough formation, I consider, is sufficient for our purpose at the moment. What is noteworthy here is that strong dispositionalists advocating epistemological dispositionalism and theorists of CI both agree that if there are categorical properties, then we are in principle ignorant of them. What differentiates the former from the latter is only that the latter has a commitment to the principle against the non-contributive (which will be explained below). This principle motivates the former not to believe in categoricalism (i.e. the existence of categorical properties).

The other assumption of the argument from ignorance is the principle against the non-contributive. There are different possible motivations for holding the principle. A notable one, which I will focus on in this section, is provided by Chris Swoyer (1982). He considers it to be an epistemic virtue that we should only posit things in our philosophical theories when they have explanatory value. For a posit to have an explanatory value, it has to be able to help us to account for some of our other knowledge. He writes:
The only reason I can see for supposing that there are such things as properties at all is that a philosophical theory of them has explanatory value, helping us to account for such things as qualitative resemblance, possibility, various semantic notions, and our ability to correctly apply predicates to novel instances. Hence the best way to learn what properties are like is to ask what something would have to be like in order to account for the things that properties are invoked to explain. (pp. 204-205, original emphasis)

An assumption Swoyer develops from the above consideration seems to be that if we cannot have knowledge of something, then it will have no explanatory value and hence should not be posited. Combining this with epistemological dispositionalism, Swoyer rejects categoricalism (strong or weak).

I shall move to the argument from science before I respond to the argument because my response to both arguments will be the same.

5.4.2. The argument from science

The term ‘argument from science’ is offered by Brian Ellis and Caroline Lierse (1994, p. 32). According to the argument, science has only discovered dispositional and causal properties. Therefore, we simply do not need categorical bases to explain our observations of the world (see, for example, Ellis and Lierse 1994, p. 32; Hawthorne 2001, pp. 368-369, Ney 2007, pp. 53-56). As Ellis and Lierse famously put it:

This is the argument from Science. With few exceptions, the most fundamental properties that we know about are all dispositional. They are of the nature of powers, capacities, and propensities. Therefore, we must either suppose that these basic properties are not truly fundamental, and that they will all eventually be shown to be dependent on categorical properties, or else we must concede that categorical realism is false. (1994, p. 32)\(^{19}\)

The assumption of the argument from science is what I call scientific dispositionalism. It is as follows:

\(^{19}\) Of course, as I mentioned before, Ellis and Lierse (1994) believe that while most dispositional properties are bare, spatiotemporal properties are irreducible to dispositional properties and are thus fundamental and categorical. This is where, I believe, the clause ‘with few exceptions’ comes from. I nevertheless choose to cite them because the cited passage is arguably the most influential version of the argument from science.
Scientific dispositionalism: All scientific knowledge is about dispositions and their manifestation (i.e. causal behaviours).

Scientific dispositionalism is strongly related to epistemological dispositionalism, and hence the argument from science is strongly related to the argument from ignorance. This is because the reason for science to have only discovered dispositional and causal properties (if this is true) is often considered to be epistemological dispositionalism. For example, John Hawthorne, a proponent of the argument from science, writes:

All scientific knowledge about negative charge is knowledge about the causal role it plays. [...] If there were a [categorical property] that were, so to speak, the role filler, it would not be something that science had any direct cognitive access to, except via the reference fixer ‘the [categorical property] that actually plays the charge role’. (2001, p. 368, my emphasis)

What he has in mind seems to have a lot of overlap with epistemological dispositionalism.

However, scientific dispositionalism and epistemological dispositionalism are still separate doctrines. While the latter entails the former, the former does not entail the latter. It might be the case that while it is epistemically possible for us to know of categorical properties, it contingently turns out to be the case that science has found no such properties. For example, Alyssa Ney simply appeals to ‘the physical scientists’ list’ of properties and suggests that categorical properties are not found there (2007, p. 54). The physical scientists’ list is not merely constrained by epistemic possibility, and thus we cannot assume that Ney is committed to epistemological dispositionalism or any similar view in epistemology.

The next step of the argument from science is a consideration of ontological economy. Its proponents argue that we should avoid committing to categorical properties because of considerations of ontological economy. The underlying reasons are diverse. Some philosophers appeal to the explanatory completeness of physical science, according to which we only need physical explanations

---

20 To be fair, Ney may not be a proponent of scientific dispositionalism, because she does not oppose the existence of all categorical properties but only the kind of categorical properties that are of concern for typical categoricalists like Lewis and Langton, i.e. categorical properties that are causally inert and fundamental and are the bases of everything else. In her view, mass and charge are intrinsic properties, so it is possible that she also considers them to be categorical. Since the way Ney categorises properties is not typical and thus requires a separate assessment, I will not go into further details of her view here.
to explain physical events (see, for example, Ney 2007, p. 55); others appeal to the principle that we should avoid positing what is not needed (see, for example, Hawthorne 2001, pp. 368-369).

5.4.3. Responding to the razor arguments

My response to the two razor arguments is that epistemological dispositionalism and scientific dispositionalism are empirically false; it is not the case that only dispositional properties have explanatory roles in science. At least some (fundamental and/or non-fundamental) categorical properties irreducible to dispositions, I will argue, do have significant and indispensable explanatory roles in science. Therefore, the razor arguments are not sound.

5.5. The binding arguments

My two arguments for categoricalism both belong to a family of arguments that I call the binding arguments. According to the arguments, categorical properties have what I call binding roles, which are roles posited to explain the commonality of some properties in microphysics. To be more specific, they are required to act as the bearer of multiple dispositional properties in order to (1) hold the spatiotemporal locations of dispositions together and (2) unify some natures of dispositions. I consider these two roles to be indispensable in current microphysics, and I will offer an argument for each of them. The arguments are what I call (I) the argument from compresence and (II) the argument from correlation. If these arguments are sound, then categoricalism is tenable, while strong dispositionalism is not. The razor arguments will also lose their force because scientific dispositionalism and epistemological dispositionalism are both false.

Both the argument from compresence and the argument from correlation are empirical because they both rest on what current empirical scientific theories are like. Since there is no evidence that natural science has reached the fundamental level, both arguments are fallible in the face of forthcoming scientific discoveries. (Of course, strong dispositionalists who appeal to the argument from science must make a similar concession.) Nevertheless, I still consider the arguments to have force. There is currently no reason to expect that what is said in the completed science will be totally discontinuous with what is said in current science. Even in different interpretations of quantum mechanics and string theory, there is, as far as I am concerned, no tendency to go against what will

---

21 The reader is reminded that the categorical properties I have been mentioning along the way are assumed to be fundamental unless specified, since they are the kind of categorical property this dissertation is interested in (see Section 4.4). Non-fundamental categorical properties irreducible to dispositions are taken into concern here because their existence in science shows that epistemological dispositionalism and scientific dispositionalism are false.
be argued in the rest of this section. Some further considerations of empirical scientific theories will be provided in Section 5.3.4.

It is also noteworthy that I consider the second argument, the argument from correlation, to be a stronger argument than the first argument, the argument from compresence. However, I shall begin with the argument from compresence because the argument from correlation follows from an assumption that I will argue for in making the argument from compresence. As the discussion goes on, the reader will see the reasoning behind this.

5.5.1. The argument from compresence

The first argument for the claim that categorical properties play the binding role is the argument from compresence. According to the argument, there are multiple dispositions that always collocate, and therefore we ought to attribute them to a common bearer object. Here I provide a hypothetical example for understanding. Let us assume that electrons, neutrons, and protons are fundamental. An electron has multiple dispositions with regard to the strong force, the weak force, the electromagnetic force, gravity, and so on. All of these dispositions always collocate and will never come apart. For example, when the disposition $w$ of the electron concerning gravity moves from position $F$ to position $G$ because of the gravitational attraction of a black hole, other dispositions $x$, $y$, $z$, and so on concerning the strong force, the weak force, and the electromagnetic force will all follow disposition $w$ to move to position $G$. The worry is that if the relation between dispositions $w$, $x$, $y$, and $z$ is nothing more than being members of the same aggregate, it will be difficult to see why they always collocate and will not come apart. What is the ‘glue’ that holds the dispositional properties together? A natural explanation for this phenomenon is that dispositions $w$, $x$, $y$, and $z$ are all dispositions held by the same object, while the object is not merely the aggregate of the dispositions.

Certainly, we all know that electrons, neutrons, and protons are not fundamental, but there is still no sign that the concept of physical objects that are the bearers of multiple dispositional properties will be eliminated. For example, a string in string theory is still being posited by science as an object bearing dispositional properties, not a single dispositional property. It is of course possible that an elimination of objects bearing dispositional properties will eventually happen in the completed science, but there is currently no reason to expect so. If we do not expect that what is said in the completed science will be totally dissimilar to and discontinuous with what is said in current science, then we have a reason to take the consideration above seriously in metaphysics.
Up to this point, categorical properties have not been mentioned. In fact, the argument above, as many readers might be reminded of, is exactly like the argument from compresence provided by substratum theorists in the debate between the substratum theory and the bundle theory. While the bundle theorist believes that an object is nothing more than a collection of properties (and the relations between them), the substratum theorist believes that an object has a substratum (i.e. a characterless substance) that bears its properties. The argument from compresence provided by substratum theorists, like the argument above, suggests that properties must have substrata in order to be bound in the same location without coming apart. The redness and the roundness of a cherry, for example, will not float out from the cherry and fly independently into the sky. The seminal formation of the argument is offered by John Locke in his *Essay Concerning Human Understanding* (1924/1689) and is advocated by his followers in contemporary analytic metaphysics, such as C. B. Martin (1980) and E. J. Lowe (2000). Locke writes:

> [B]ecause we cannot conceive how [qualities] should subsist alone, nor one in another, we suppose them existing in and supported by some common subject; which support we denote by the name substance, though it be certain we have no clear or distinct idea of that thing we suppose a support. (ch. 2.23.4)

Categorical properties are not substrata, but I will argue that the same argument can be applied to the existence of categorical properties because I believe that categorical properties can explain compresence just as substrata do. In addition, I will argue that the weakness of the substratum theory will not be found if what are of concern are categorical properties, and hence categoricalism might be a better explanation for compresence than the substratum theory.

In the case of the hypothetical fundamental electron, for example, the electron, as an object, acts as the bearer of all of its dispositions. It is distinct from them because otherwise the relation between the bearer and the dispositions it bears should be described as an identity relation but not as a bearing relation. (I will return to a version of a theory of the former in Section 5.6.1.) A difficulty of the substratum theory is that the notion of a substratum seems to be incoherent and thus implausible. On the one hand, a substratum is a substance with all of its properties abstracted away. On the other hand,

---

22 It is important to note that the reading of Locke I present here is the mainstream interpretation by contemporary analytic metaphysicians working on the substratum theory and the bundle theory, such as Martin (1980) and Lowe (2000). Whether this interpretation accurately represents Locke’s actual view or not is a difficult question in Locke scholarship, which I will not address here.
it is supposed to be a real entity. With these two presuppositions combined we have a ‘real empty thing’, which both exists and has no property. The idea is difficult to understand. For example, to say that there is a ‘pure electron’ or an ‘empty electron’ without any properties of an electron seems to be implausible. Locke admits that he has no idea what a substratum is. He writes:

[T]hat is the idea of substance, which we neither have nor can have by sensation or reflection. (1924/1689, ch. 1.4.8)

[We] signify nothing by the word substance but only an uncertain supposition of we know not what, i.e. of something whereof we have no (particular distinct positive) idea, which we take to be the substratum, or support, of those ideas we do know. (1924/1689, ch. 1.4.8)

[O]f substance, we have no idea of what it is, but only a confused, obscure one of what it does. (1924/1689, ch. 2.13.19)

Certainly, we have seen that many categoricalists also believe that we are irremediably ignorant of categorical properties. However, the worry is not that we are irremediably ignorant of the substratum, but that the irremediable ignorance of the substratum reflects the incoherence of the concept of a real empty thing.

Despite this, if what the argument is arguing for is the existence of categorical properties, then there is no commitment to the idiosyncratic concept of substratum. We simply need to suppose that there is an extra property that is the bearer of the dispositional properties. Even if it is considered unknowable by many, being unknowable is not the same as being incoherent. It is, after all, difficult to see why being a bearer object cannot be considered a special kind of property. We can simply posit the bearer object of all of the dispositions of the hypothetical fundamental electron (we discussed earlier) as a property. The property of being an electron can be considered different from the sum of the dispositions possessed by the electron.

Furthermore, the argument from compresence for categorical properties can resist the major objection to the argument from compresence for substratum provided by bundle theorists. Many bundle theorists, in response to the argument from compresence for substratum, suggest that there is a primitive and inexplicable relation that acts as a tie to bind the properties together (see, for example, van Cleve 1985, pp. 97-98; Denkel 1992, p. 710). The relation, in their view, is not a substratum. This response, like the substratum theory, is not without weakness because the relation is inexplicable just as substrata are. Jiri Benovsky notes that the bundle theory and the substratum theory are actually
very similar in positing inexplicable entities to explain the binding of properties, and thus there may be no real disagreement between them (2008, pp. 177-178).

The argument from compresence for categorical properties can provide a new way to deal with the controversy. If one follows the bundle theorist and posits a primitive and inexplicable relation that binds the dispositions, she has already posited a relational property. It seems to me that the property can be best understood as a categorical property that bears the dispositions. This understanding cannot be resisted by the principle of simplicity, which many strong dispositionalists consider to be a reason to deny the existence of categorical properties, because nothing is changed in the ontological economy. With the assumption that a relation has to be posited, what I provide is merely an understanding of the relational property as a categorical property; no new property is added. (With this understanding of compresence, it can be said that my view is a version of the bundle theory – though perhaps not the mainstream version – because an object is, in my view, nothing more than a bundle of properties, and there are no additional substrata that bear them, and this exactly matches the definition of the bundle theory. Nevertheless, since the bundle theory is not the interest of this section, I shall not go into any further detail. But I hope that my disagreement with the substratum theory is clearer when my sympathy with the bundle theory is explicit.)

Can we, instead, understand the relational property to be a dispositional property? Not impossible, but very idiosyncratic and ad hoc. Following this view, when an electron moves from position $F$ to position $G$ because of the gravitational attraction, other non-gravitational dispositions have to be considered as being ‘pulled’ to position $G$ by the manifesting gravitational disposition. Science tells us nothing about such idiosyncratic pulling force. It simply tells us that there is an electron with dispositional properties $w, x, y,$ and $z$, and it moves from position $F$ to position $G$. The categoricalist can make sense of the scientific theories by understanding the electron as a categorical property, thereby having a categorical property for free. But the strong dispositionalist has to disagree with the scientist and posit an extra ‘force’ that does the pulling of dispositions posited by science. She not only fails to perform better than the categoricalist in terms of simplicity and staying close to science as some proponents of the razor arguments claim, she actually needs to disagree with the literal understanding of the scientific theories and thereby fails to have her desired respect for

---

23 No doubt, intuitively, an electron is an object with properties, not a property. But I do not consider this to be a cost of my view, at least not a significant one, because I do not believe that our intuition about what an object and a property really are is very precise. In my view, the view that a fundamental object might be called ‘a fundamental property’ does not seem to be counterintuitive, and vice versa.
science. (Of course, there are some alternative hypotheses available to the strong dispositionalist. I will discuss them in Section 5.6.)

Up to this point, the argument is developed from an analogy to the argument from compresence for the substratum theory, and the categorical properties in question are simply posited as ‘being the bearer objects’. This kind of property might be too abstract for some readers to consider as a genuine property. Some might simply prefer the substratum theory, or believe that the view I put forward has no real difference from it. Nonetheless, as I have mentioned, there is another (and, in my view, stronger) binding argument apart from the argument from compresence, which is the argument from correlation. With the two binding arguments combined, it will not only be shown that the consideration of binding roles of categorical properties is different from that of substratum (or the inexplicable relation posited by some bundle theorists), it will also provide us with a much stronger motivation to believe in the existence of categorical properties. In particular, it will be shown that categorical properties do have features and are not featureless like substrata by definition are. To put this in David Chalmers words, what I am advocating is a ‘thick quiddity picture’ rather than a ‘thin quiddity picture’ (2012, pp. 349-350).

5.5.2. The argument from correlation

The argument from correlation, much like the argument from compresence, appeals to the commonality between some dispositions and supports the claim that they share a common bearer. What differentiates it from the argument from compresence is that it appeals to the common quantitative features involved in the dispositions, such as mass, charge, size, and so forth, but not the locations of the dispositions. This makes the argument stronger than the argument from compresence and can also be seen as a supplement to it. This is because, even though it is also an appeal to the need of bearer objects, it not only suggests that there are bearer objects, but also shows that there are some properties intrinsic to them, which are not dispositional.

Let us return to the hypothetical fundamental electron. It has multiple dispositions with regard to the strong force, the weak force, the electromagnetic force, gravity, and so on. What scientific theories have shown is that these dispositions all correlate with the same set of quantitative features, such as mass, charge, size, and so on. If there are only dispositions but no categorical properties, the quantities have to be understood in terms of the nature of each disposition. Let us take mass, for example. The mass $m$ of the electron has to be understood as the quantity of its disposition to behave under gravitational attraction, and the quantity of its disposition to behave under the strong force, and
so on. We then have mass properties \( m_1, m_2, m_3, \ldots, m_n \) with regard to each disposition, say, a mass property \( m_1 \) with regard to gravitational attraction, a mass property \( m_2 \) with regard to gravitational attraction, and so forth; and, as we all know, they all have the same quantity. Two questions will arise. First, is there anything that can explain the fact that \( m_1, m_2, m_3, \ldots, m_n \) have the same quantity? Secondly, can we consider \( m_1, m_2, m_3, \ldots, m_n \) to be identical but not numerically the same? It appears that strong dispositionalism leads us to answer ‘no’ to these questions. This implies that \( m_1, m_2, m_3, \ldots, m_n \), contrary to what science says, are in fact independent values. In addition, their same quantity can only be understood as mere coincidence. This seems to be very implausible. (Of course, this is contestable, since there are some more sophisticated hypotheses available to the strong dispositionalist, which I will discuss in Section 5.6.)

Meanwhile, the categoricalist can simply explain \( m_1, m_2, m_3, \ldots, m_n \) as a single categorical property of the bearer object that bears all of its dispositions.\(^{24}\) No doubt, it can be questioned why quantitative features such as mass and size would collocate (see, for a similar argument, Martin 1980, p. 4), but the categoricalist has no difficulties in responding to this objection. She can simply understand them as non-fundamental categorical properties that are realised by or supervene on the fundamental categorical property of the electron, which is a singularity. In comparison with the strong dispositionalist who (in the absence of the possible responses I will discuss in Section 5.6) has to understand the electron as having many independent mass properties, the categoricalist who posits a single non-fundamental categorical property that takes the roles of \( m_1, m_2, m_3, \ldots, m_n \) obviously has an advantage in matching with scientific theories. This is because what our current scientific theories do is simply consider \( m_1, m_2, m_3, \ldots, m_n \) as a single attribute of the object that bears the dispositions, not as features of independent dispositions. While the categoricalist understanding of the quantitative features can take the scientist’s claims at – or, moderately speaking, closer to – face value, the strong dispositionalist seems to have difficulty in doing so.

A combination of the argument from compresence and the argument from correlation can allow us to see why fundamental categorical properties, in my view, are not substrata (as some might believe). Accordingly, fundamental categorical properties have natures or features that realise non-fundamental categorical properties like mass, charge, size, and so forth. They are thus different from substrata, which are by definition featureless.

\(^{24}\) Most theorists, categoricals included, believe that mass and charge are dispositional properties (see, for example, Blackburn 1990), but I believe that this assumption can be challenged. For examples of authors who share similar views to mine, see Smart 1963; Ney 2015.
5.5.3. Science and explanations

As I have shown, microphysics posits (1) roles that are required to hold the spatiotemporal locations of dispositions together and (2) roles that are required to unify the natures of dispositions, which are, respectively, (i) objects and (ii) some quantitative features that are intrinsic to the objects and have correlations with multiple dispositions of the objects. It is difficult to see how these roles could be eliminated from current scientific theories. I have argued that neither is reducible to dispositions and that both should be considered (fundamental or non-fundamental) categorical properties. This suggests that (fundamental or non-fundamental) categorical properties have indispensable roles in scientific explanations. If this is correct, the place of categorical properties in science is now settled. The assumption of the razor arguments, epistemological or scientific dispositionalism, is false. Furthermore, strong dispositionalism is untenable. It would then be useful to consider the reason why some philosophers are committed to epistemological or scientific dispositionalism and thus strong dispositionalism.

My hypothesis is as follows. It seems to me that it is tempting to commit to epistemological or scientific dispositionalism because many only have causal explanations in mind when they consider the nature of explanations, but this is incorrect because causal explanations are not the only kind of explanation available.

In a causal explanation, we trace back through history to search for the cause(s) of the phenomena to be explained. Consider the death of John Kennedy. The event can be explained by the event of a bullet flying towards him, the event of a gun firing, the event of the assassin firing his/her gun, and so forth. All of events are diachronically and causally related to the death of John Kennedy. If causal explanations are assumed to be the only kind of explanation available, it will be tempting to think that an explanation can only posit entities that have a role in the causal chain.

However, the thought that all explanations are causal explanations is simply false. In science, we also have structural explanations, in which we explain a phenomenon by the underlying structure that realises it. Consider the neuroscientific study of near-death experience (NDE), which many people consider to be an experience of the afterlife. Some neuroscientists suggest that NDEs are generated by the same brain mechanisms that generate lucid dreams and are thus, like lucid dreams, hallucinations. The explanation is not diachronically related to the phenomenon to be explained and is thus not a causal one, because the brain mechanisms that generate NDEs and lucid dreams do not

---

25 Of course, both the phenomenon to be explained and the phenomenon used in the explanation have causal elements.
cause them but realise them; the relations between the brain mechanisms and these events are not
diachronic but synchronic. Another interesting example of structural explanation is some versions of
the multiverse theory in physics, which provides a clear case that explanations can be provided when
causal connections are absent. According to these theories (and their advocates), our universe is not
causally connected to some other universes, but we can still know of their existence. This is because
our knowledge of the causal structure of our universe motivates us to best explain it as a part of the
multiverse, which includes some other universes as its parts. In sum, there is no reason to assume that
entities posited by structural explanations must have a role in the causal chain.

The two binding roles I attribute to categorical properties are both posited by (or as) structural
explanations. They are the underlying realisers of dispositions that tell us why the dispositions are
structured in the way we observe but are not part of the diachronic story of how the dispositions are
cause by entities or properties involved in prior events.

5.6. Alternative hypotheses available to the strong dispositionalist

In Section 5.5, I argued that categorical properties have binding roles in scientific explanations by
appeal to the argument from compresence and the argument from correlation. In this subsection, I
will consider two alternative hypotheses available to the strong dispositionalist: (1) the single
disposition hypothesis and (2) the non-causal and unconditional disposition hypothesis.

5.6.1. The single disposition hypothesis

According to the single disposition hypothesis, a fundamental object consists of (or is identical to) a
single dispositional property, not a collection of multiple dispositional properties. Let us return to the
hypothetical scenario that electrons, neutrons, and protons are fundamental. Being the electron role
(or the realiser of it) will be considered by the view to be a fundamental dispositional property. Its
dispositions with regard to the strong force, the weak force, the electromagnetic force, gravity, and
so on are not what constitute it but the second-order features derivative of it. In this case, there are no
multiple dispositional properties in need of binding.

My response to the single disposition hypothesis is as follows. First of all, even if the hypothesis
can settle the argument from colocation, the argument from correlation seems to remain unanswered.

The explanation is not a causal explanation, because the two phenomena are not causally related, not because they are not
causal in nature.
A fundamental dispositional property is now considered a complex, if not idiosyncratically complex, conjunction of dispositional features following different laws, but within this conjunction there is still nothing that explains why these dispositional features share an identical set of quantitative features, such as mass, charge, size, and so forth. As we have seen, the categoricalist can simply avoid the difficulty by saying that a categorical property possesses multiple dispositional properties, and the multiple dispositional properties correlate to the same set of quantitative features because the quantitative features are non-fundamental categorical properties that are solely realised by the categorical property. However, it seems that what the proponent of the single disposition hypothesis can answer is, on the other hand, only that it is a brute fact that cannot be explained by any further fact. This is unsatisfactory. Of course, it is possible that she can solve the problem by offering more additional hypotheses, but it is at least the case that the problem cannot be solved by the hypothesis solely.

There is another, and perhaps bigger, worry with the single disposition hypothesis. The worry is that the view does not allow us to categorise dispositional properties of different fundamental objects into the same first-order kind, and some radical consequences will follow. Let us return to the hypothetical scenario that electrons, neutrons, and protons are fundamental. We can expect that scientists would consider the gravitational forces of electrons, neutrons, and protons to be all members of the same kind, a fundamental force named gravity. However, the single disposition hypothesis cannot allow gravity to be a fundamental force. This is because the fundamental forces in this case will be the fundamental dispositional properties of an electron, a neutron, and a proton representatively, and that gravity is a derivative and second-order feature that we derive from those fundamental dispositional properties. In addition, an electron, a neutron, and a proton will not share any fundamental force. These consequences are radically counterintuitive and against ordinary scientific assumptions.

We have seen that the single disposition hypothesis leads us to commit to a metaphysical framework that is unsatisfactory, counterintuitive, and against ordinary scientific assumptions. While many strong dispositionalists believe in strong dispositionalism and reject categoricalism because they believe they are following the lead of science, it is difficult to see how the single disposition hypothesis can be compatible with that attitude.
The views that some dispositions are non-causal and/or unconditional in nature, though not widely discussed, has recently gained some attention. These views might provide some ways for the strong dispositionalists to respond to the binding arguments. Before I introduce the non-causal and unconditional disposition hypothesis, I should firstly introduce the views.

Let us begin with non-causal dispositions. As far as I am concerned, there is only one article on the topic (at the time of writing this dissertation), which is Daniel Nolan’s recent article ‘Noncausal dispositions’ (2015). In the article, Nolan notes that dispositions are often assumed by many, categoricalists and dispositionalists included, to have an ‘intimate connection’ to causation (p. 425). But this assumption, he believes, might be false because there might be some dispositions that are non-causal dispositions. He argues that if this is true, many standard accounts of categorical and dispositional properties might be in need of modification. While this dissertation follows the standard assumption in the literature that dispositions and causal powers are the same, Nolan’s warning is worth considering.

The kinds of dispositions Nolan considers are those in the form ‘X is disposed to Φ if C’, where Φ is a manifestation and C is a condition (p. 425). Non-causal dispositions, Nolan explains, are dispositions where the relations between Φ and C are not causal (p. 426). He gives several examples of dispositions that are possibly non-causal, even though he makes it clear that it is ‘open to a theorist’ to understand them otherwise (p. 427). One of the examples provided by Nolan is the EPR experiment in quantum physics. In the experiment, the chance of a measurement turning out a certain way depends on whether we have the information of a former experiment. Many scientists consider the relation between the first and the second experiments to be non-causal because they can be set in a way that only a signal faster than the speed of light, which is impossible according to the theory of relativity, can to be sent from the first to the second.

The other kind of disposition that casts doubt on the common assumption is the unconditional disposition. Dispositions are typically considered to be conditional (see, for example, Armstrong 1997; Smith & Stoljar 1998), but some recent studies suggest that there are unconditional dispositions. In this case, the clause ‘if C’ in Nolan’s form of disposition ‘X is disposed to Φ if C’ has to be removed. One example that gains attention is that of a radioactive atom’s disposition to decay, which is unaffected by any known process (see, for example, Childers 2013, p. 40).
The reason I introduce the concepts of non-causal dispositions and unconditional dispositions is that a combination of them, which is the concept of non-causal and unconditional disposition, might provide an alternative explanation of the bindings I considered. The idea is that there are some non-causal and unconditional dispositions that bind the causal dispositions we observe. Rather than generating causal behaviours under some particular conditions, these non-causal and unconditional dispositions unconditionally ‘affect’ some features of the causal dispositions we observe. For example, the mass of an electron might be considered as a non-causal and unconditional disposition that ‘affects’ some quantitate features of its causal dispositions (including those associated with gravitational force, electromagnetic force, strong force, weak force, and so forth).

I consider this possible hypothesis to be more plausible than the single disposition hypothesis, but it also has a number of problems. Firstly, even if some common quantitate features of causal dispositions like mass and charge can be understood as non-causal and unconditional dispositions, what can the hypothesis tell us about the argument from colocation? Certainly it can be argued that there are non-causal and unconditional dispositions that generate the colocation of a set of causal dispositions, and this, I assume, is more plausible than saying that a causal disposition does so because the latter seems to, in turn, imply that there is an extra ‘force’ that does so, which is idiosyncratic and against our scientific assumptions. But the fact that the view is less idiosyncratic and against our scientific assumptions does not imply that it is not so. As I have argued, in science it is considered to be a fact that causal dispositions collocate simply because they belong to the same object, and there are no suggestions of any further connection that connects them. To argue that a non-causal and unconditional disposition does so still seems to be difficult to match well with what science tells us.

Secondly, recall that the argument from colocation can be applied to common quantitate features of (causal) dispositions, just as they can be applied to (causal) dispositions. If we want to dispense with categorical properties, we cannot simply understand the common quantitate features of causal dispositions as non-causal dispositions; we also have to consider that there is an additional non-causal disposition that fixes other non-causal dispositions into the same location. Let us compare this hypothesis with the categoricalist hypothesis that the common quantitate features of dispositions are non-fundamental categorical properties realised by a fundamental categorical property. The use of the conception of realisation here seems to have made the view simpler than the view of the strong dispositionalist; by the conception of realisation, only one fundamental categorical property is posited.

---

26 I add quotation marks here because the term ‘affect’ is understood by many to be causal, which is not the case here.
as the basis of a fundamental object, while the strong dispositionalist has to consider a fundamental object to be a complex network of fundamental dispositions – which includes its causal dispositions, its non-causal dispositions that regulate some quantitate features of its causal-dispositions, and the non-causal disposition that fixes other non-causal dispositions into the same location. It is unclear that the strong dispositionalist committing to the view has the upper hand over the categoricalist in terms of theoretical virtues (such as simplicity, explanatory power, predictive power, seminality in suggesting future work, and so on), among which simplicity is considered by many theorists, myself included, to be a very important one.

Thirdly, the existence of non-causal dispositions (if there are any) is damaging to both current categoricalist and strong dispositionalist accounts of dispositional properties because most of them assume that all dispositions are causal in nature. The writers making this assumption pointed out by Nolan include influential categoricalist writers such as Lewis and Jackson and influential strong dispositionalist writers such as Bird (Nolan 2015, p. 425). And, in fact, many other influential strong dispositionalist writers such as Shoemaker (1980) and Swoyer (1982) make the same assumption. This creates problems for many arguments for each doctrine, such as the razor arguments. As we have seen, many strong dispositionalists commit to epistemological or scientific dispositionalism in their razor arguments for strong dispositionalism, and the reason for believing in epistemological or scientific dispositionalism is often the view that we can only know things through being causally connected to them. If it is true that we can know of some non-causal dispositions, then the reason many believe in epistemological or scientific dispositionalism will be undercut.

In addition, so far I am following the majority assumption that categorical properties are defined by being properties that are not causal or dispositional, while the dispositional properties here are considered to be causal powers. However, if there are non-causal dispositions, then we might have to reconsider the definitions of categorical and dispositional properties. With a revision of the definitions, perhaps what typical strong categoricalists such as Armstrong, Lewis, and Langton consider to be fundamental categorical properties can all be (non-causally) dispositional without being inconsistent with what these writers intend to mean. In sum, it is, currently, difficult to see how we should understand the concepts of categorical and dispositional properties if we take non-causal dispositions into account.

A serious and independent inquiry into the implications of the existence of non-causal dispositions (if there are any) has to be carried out, but there are, as far as I am concerned, not many studies in the literature (at the time of writing). I am unable to offer a deeper inquiry into the topic.
because of the limit of scope; and it cannot be assumed that the existence of non-causal dispositions will be beneficial to the strong dispositionalist. With this in mind, for now it might be best to set non-causal dispositions aside and consider categoricalism (strong or weak) to be a currently more promising hypothesis than strong dispositionalism. Whether and how considerations of non-causal dispositions can motivate changes to this assumption can be left for future study.

5.7. Some further considerations of empirical science

I will end this section by considering empirical scientific theories, focusing on ontic structural realism (OSR) advocated by James Ladyman and Don Ross. Up to this point, for the sake of having a simpler point of departure, most of my discussions of categoricalism and strong dispositionalism have purposefully made the assumption that the traditional framework of metaphysics is correct. According to the framework, space is a mere background or container and there are only atomistic space-occupiers interacting within. Whenever I mentioned categorical properties earlier in this dissertation, I made the assumption that they are such atomistic space-occupiers, in the sense that each categorical property is like a fundamental atom. Unfortunately, the framework is challenged by empirical scientific discoveries. Since the discussion in this section rests on the value of categorical properties in scientific explanations, it is important to move to empirical scientific theories and assess whether the view I advocate is really compatible with them.

The strongest case against categoricalism that appeals to current scientific discoveries is probably provided by Ladyman and Ross’s ontic structural realism (OSR). I will assess the view and spell out its implications for the debate.

5.7.1. Ontic structural realism (OSR)

In their seminal book *Everything must go* (2007), Ladyman and Ross argue their *ontic structural realism* (OSR) is a theoretical framework better than (Langton, Lewis, and Jackson’s versions of) categoricalism and should simply replace it. OSR is the view that the relational structure of the world is ontologically fundamental, but is not one that consists of individuals with intrinsic properties. Identity and individuality of objects, in their view, depend on the relational structure.

Ladyman and Ross develop OSR from an analysis of empirical science, especially quantum physics where quantum particles are found not to have exact space-time locations. In their view, quantum particles should be given a non-individualistic interpretation, in which concepts of individual objects should be eliminated. Accordingly, our concepts of individual objects are
unsupported in the context of field theory (p. 140). We have these concepts simply because of
distinguishability or discernibility, not objective individuality (p. 134). Consequentially, they list the
standard assumptions in metaphysics below and argue that they are all challenged by OSR:

(i) There are individuals in spacetime whose existence is independent of each other. Facts
about the identity and diversity of these individuals are determined independently of their
relations to each other.
(ii) Each has some properties that are intrinsic to it.
(iii) The relations between individuals other than their spatio-temporal relations supervene
on the intrinsic properties of the relata (Humean supervenience).
(iv) [The Principle of the Identity of Indiscernibles (PII)] is true, so there are some properties
(perhaps including spatio-temporal properties) that distinguish each thing from every other
thing, and the identity and individuality of physical objects can be accounted for in purely
qualitative terms. (p. 151)

For Ladyman and Ross, Lewis is merely one of the traditional metaphysicians who is already
proved wrong by the latest scientific discoveries. This is especially obvious when Lewis advocates a
version of categoricalism that he calls ‘Humean supervenience’. As Ladyman and Ross cite Lewis:

[A]ll there is to the world is a vast mosaic of local matters of particular fact, just one little
thing and then another. […] We have geometry: a system of external relations of spatio-
temporal distance between points (of spacetime, point matter, aether or fields or both). And
at these points we have local qualities: perfectly natural intrinsic properties which need
nothing bigger than a point at which to be instantiated. […] All else supervenes on that. (p.
148)

Unsurprising, they simply consider the doctrine to be false because science has turned out to say
otherwise. It assumes the existence of individuals with intrinsic natures, and this assumption has
already been rejected by our best physics (p. 154). They also consider other versions of categoricalism
like that of Langton to be false, because they make the same mistake Lewis makes.
5.7.2. Implications of OSR

I believe that OSR is more attractive than the standard framework of traditional metaphysics because the view Ladyman and Ross propose seems to have pointed out something correct in quantum physics. In the past decades, a growing number of philosophers are taking the theory of quantum nonlocality more seriously, according to which particles are, in fact, not space-time points and even quantum holism, according to which particles are, in fact, not separate entities; and they warn us that it seems difficult to avoid the challenges these doctrines provide to the traditional view of atomistic space-occupiers held by most analytic metaphysicians (see, for examples of relevant discussions, Teller 1986; Oppy 2000; Schaffer 2003; Esfeld 2004; Papineau 2008). In fact, even Lewis himself admits that quantum physics is incompatible with his version of categoricalism, and he is ‘ready to believe’ in quantum physics. He simply rejects the science because he ‘must see how it looks when it is purified of instrumentalist frivolity, and dares to say something not just about pointer readings but about the constitution of the world’ (1986a, p. xi). OSR’s attempt to revise the assumptions of metaphysics, I believe, is what the inquiry into the nature of categorical properties and dispositional properties should take seriously. We should not assume there are individual ordinary objects simply because of our intuitions and everyday experience. As Frank Jackson famously puts it, ‘Given a choice between going with science and going with intuitions, go with science’ (2003, p. 251). Science has shown that many of our intuitions and everyday assumptions are mistaken. The philosopher should weight science more than our intuitions and everyday assumptions when she is looking for input data of her philosophical reflection.

However, does OSR really refute categoricalism as Ladyman and Ross suggest? I believe that the answer is no. What Ladyman and Ross have learned from our best science is that our universe is a single relational structure. But this ontological holism, as I understand it, does not seem to be incompatible with categoricalism. It seems that a combination of the two views is possible, according to which the single relational structure described by OSR has a categorical base. What follows seems to be a form of monism, according to which there is only one single fundamental categorical basis that realises everything. Despite its counterintuitiveness, there has been a recent a revival of monism in metaphysics, and the view has attracted followers (see, for example, Schaffer 2010; Schroer 2010b).

Certainty, it might be questioned whether the binding arguments are compatible with OSR, even if the version of OSR being considered is the monistic categoricalist version. This is because OSR denies the existence of individual objects. Nevertheless, as I have argued, bearer objects and quantitative features intrinsic to them are still indispensable from science. There is no need to assume...
that the bearer objects are really (strictly) individuals or atoms. They may be considered non-
fundamental categorical properties that supervene on the fundamental monistic categorical basis.

However, I do not wish to commit to monism. My purpose here is simply to show that OSR is not inconsistent with the arguments I have defended in this section. The reasons for the purpose are as follows. Firstly, there are many difficult theoretical concerns in philosophy of physics and physics itself that have to be considered before OSR is assumed true, and I must admit that the consideration of physics above is very limited. For example, in theoretical physics, there are more than ten interpretations of quantum physics, and it is currently difficult to see which one is true (or best). An inquiry into these matters cannot be carried out here. After all, deciding whether OSR is correct must take a comprehensive consideration of our best physics into account, but this is beyond the scope of this dissertation.

Secondly, OSR and monism, and the views in philosophy of physics and physics they rely on them, are still controversial – and often minority – positions. For the purpose of this dissertation, it will be useful to purposefully leave these controversies unsettled. As I have mentioned before, the purpose of this dissertation to assess whether contemporary metaphysical naturalism, which many contemporary philosophers align themselves with, can actually be justified when it faces the challenges followed from the categorical ignorance thesis (CI). With fewer assumptions about the nature of categorical properties, the assessment provided by this dissertation can be appealing to more philosophers who share the alignment. Therefore, even though I find monism quite attractive, I will remain neutral on whether it is true in the rest of this dissertation.

5.8. Conclusion

The razor arguments offered by the strong dispositionalists suggest that categorical properties have no place in science and in explanations and thus should not be taken to exist. In response to the razor arguments, I have provided two binding arguments for categoricalism (strong or weak), which motivate us to believe in not only the existence of categorical properties, but also their explanatory value in science. This also shows that the assumption of the razor arguments, epistemological or scientific dispositionalism, is false and that, furthermore, strong dispositionalism is untenable.

As I have said, the debate between categoricalism and strong dispositionalism is an important focus of research in the literature. It is impossible to discuss all major arguments for and against each doctrine in a single section. Therefore, my assessment of the doctrines is not conclusive, because there might be other independent arguments for strong dispositionalism or against categoricalism in
the literature that can outweigh the evidence for categoricalism outlined here. And, certainly, there might also be other independent arguments for categoricalism that can outweigh their evidential force. Unfortunately, I can only assume that all independent arguments have no force, because of the limit of scope. However, if my arguments are successful, we at least have sufficient motivation to assume that there are categorical properties, which are the main character of the categorical ignorance thesis (CI). This assumption allows us to discuss CI in the rest of this dissertation.
Part II

Section 6. The nature of categorical properties

6.1. Introduction

In Section 5, I provided some motivations to believe in the existence of categorical properties. Before moving to Section 7, which is an assessment of arguments for CI, I shall discuss a debate on the nature of categorical properties, which is between strong and weak categoricalism. This is because many arguments for CI (as we shall see in Section 8) are dependent on or assume particular positions within the debate. With this in mind, it would be useful to assess the debate before we discuss CI.

The readers are reminded that strong and weak categoricalism are as follows:

**Strong categoricalism:** All fundamental properties are categorical properties without dispositional essences intrinsic to them. For there to be dispositional properties, categorical properties are required to act as their bases. But the relation between the two kinds of properties (i.e., which categorical property is playing which dispositional role) is determined by the laws of nature – or what described by the laws if the laws are understood to be descriptions – that are extrinsic to categorical properties, not the intrinsic essences of categorical properties.

**Weak categoricalism:** Dispositional properties are roles played categorical properties, or are identical to the categorical properties that bears them. In the former case, the nature of a dispositional role of a categorical property is determined by the intrinsic dispositional essences of the property. The laws of nature, in the former case, are (the descriptions or analyses of) the essences, and in the latter case, are the descriptions or analyses of the properties.

---

27 This section is developed from a manuscript of an unpublished article I co-authored with Andrew James Latham and David Braddon-Mitchell, ‘The argument from modal intuition against dispositionalism: a critique’, in which I take the role of the first author.
The reader is also reminded that by a dispositional essence, I mean any feature or nature of a property that determines its dispositional role. For examples of accounts of extrinsic laws of nature, see Section 5.2.

In Section 5, when I argued for the existence of categorical properties, the arguments I used implied a view on the nature of categorical properties. More precisely, the view aligns with what David Chalmers (2012, pp. 349-350) calls a *thick quiddity picture*, according to which categorical properties do have natures or features, as opposed to being featureless. Nevertheless, I will argue that the picture implies neither strong nor weak categoricalism. In Sections 6.2, I will defend this claim. Then, in Section 6.3, I will defend my expectation that the debate between strong and weak categoricalism probably has to be settled by taking our best physical theories into account; and I will argue that it suits this dissertation best to remain neutral in the debate. In Section 6.3, I will examine an important and influential philosophical argument for strong categoricalism provided by Langton and Lewis, which is the *argument from modal intuition*. The argument points out an intuition that things can have causal behaviours different from their actual ones (in other possible worlds) and argues that, given that the intuition is correct, weak categoricalism is false. More details will be provided in the subsection. If the argument is correct, strong categoricalism should be straightforwardly considered as true, and weak categoricalism should be straightforwardly considered as false from the armchair regardless of empirical considerations. However, I will show that the argument is not sound and hence cannot threaten my decision to suspend judgment concerning whether strong or weak categoricalism is true.

6.2. The issue of my thick quiddity picture

David Chalmers distinguishes between what he calls a ‘thick quiddity picture’ and a ‘thin quiddity picture’ (2012, pp. 349-350). The thick quiddity picture is the view that categorical properties do have natures or features; the thin quiddity picture is the view that categorical properties are featureless, with the only exception of having numerical identities. In Section 5.5, I argued that a (fundamental) categorical property does have features or natures that realise the non-fundamental categorical properties of a fundamental object like mass, charge, size, and so forth. Accordingly, my view aligns with the thick quiddity picture.\(^{28}\)

\(^{28}\) The reader is reminded that the categorical properties I have been mentioning along the way are assumed to be fundamental unless specified, since they are the kind of categorical property this dissertation is interested in (see Section 4.4).
Some might believe that my view seems to be a version of weak categoricalism. The reason is as follows. The kind of non-fundamental categorical property I mentioned is factors determining how things causally behave, and hence these properties seem to be part of the laws of nature. This seems to imply that weak categoricalism is correct that the laws of nature are intrinsic to categorical properties (see the definition of the doctrine in Section 6.1).

While I believe that my view is consistent with weak categoricalism, I disagree that the former implies the latter. Instead, I believe that my view is compatible with strong categoricalism as well. This is because the non-fundamental categorical properties alone do not logically necessitate dispositions. For example, let us assume that electrons are fundamental. An electron’s mass property correlates with its dispositions, such as those with regard to the strong force, the weak force, the electromagnetic force, gravity, and so on. It seems to be a reasonable view that, in another possible world, the electron – or the categorical property playing the electron role – could be causally inert or not share some dispositions of its actual counterpart, even though it bears the same set of non-fundamental categorical properties like mass, charge, and size (or features or natures that realise the non-fundamental categorical properties).

The way to put these together is that disposition roles are extrinsically given to categorical properties, with some of their features influenced or determined by the non-fundamental categorical properties that the categorical properties bear. For example, the gravitational laws that a fundamental particle behaves according to extrinsically govern the particle or the categorical property that plays the role of the particle. But the gravitational laws do not offer the particle the quantity that takes the role of the mass property involved in the gravitational behaviours; instead, they leave the quantity to be set according the nature of a particular non-fundamental categorical property of the particle. This model allows strong categoricalism and my thick quiddity picture to compatible with each other.

I should make it clear that I am not advocating the model I provided above, but merely suggesting that it is a possibility that my thick quiddity picture and strong categoricalism can be compatible with each other. On the other hand, I believe that my thick quiddity picture is consistent with weak categoricalism as well. This is because the dispositional essences of a categorical property might be intrinsic to the property, just as its features or natures that realise the non-fundamental categorical properties it bears are.
6.3. My position in the debate between strong and weak categoricalism

There are many arguments in the literature in the debate between strong and weak categoricalism, but I believe that acknowledging the uncertainty involved in the debate and suspending judgment is the best position for this dissertation. I will defend this position in this subsection. This is an important step of this dissertation despite the fact that what will be defended is neutrality but not any substantive view. This is because, as I mentioned before, many arguments for CI (as we shall see in Section 7) are dependent on or assume strong and weak categoricalism. My defence of suspending judgment means that I do not take such arguments for CI to be compelling, and in addition, is a defence of my approach to CI in Section 7, which, unlike many theorists of CI, does not presuppose strong or weak categoricalism.

I am pessimistic that the debate between strong or weak categoricalism can be settled from the armchair. The reason for my pessimism is that determining whether dispositional essences or the laws of nature are intrinsic or extrinsic to categorical properties requires us to have an understanding of our empirical science, because categorical and dispositional properties are both parts of the physical world. As I have argued in Section 5.7.2, both the traditional metaphysical framework of philosophy and our everyday intuitions about what the world is like are challenged by what current scientific theories have shown us.

Unfortunately, although in Section 5 I put forward my thick quiddity picture as one that is based on empirical considerations, in Section 6.2 I have shown that it does not tell us whether strong or weak categoricalism is true. With this in mind, it seems that much more consideration of our best physics and philosophy of physics have to be taken into account. James Ladyman and Don Ross’s ontic structural realism (OSR), which I have discussed in Section 5.7, provides an interesting example of how such an account can be possible.

Nevertheless, as I have argued in Section 5.7, I believe that a sophisticated discussion that takes our best physics and/or philosophy of physics into account has to be left for other researches and cannot be done in this dissertation. This is because, current physical theories are too complex to be within the scope of this dissertation. Furthermore, I have also argued in the section that it will be useful to purposefully leave such considerations unsettled for the purpose of this dissertation. This is because views in physics and philosophy of physics often involve controversial and minority positions. The purpose of this dissertation is to assess whether contemporary metaphysical naturalism, which many contemporary philosophers align themselves with, can actually be justified when it faces
the challenges that follow from the categorical ignorance thesis (CI). With fewer assumptions about
the nature of categorical properties, the assessment provided by this dissertation can be appealing to
more philosophers who share the alignment. In sum, I believe that it suits this dissertation best to
remain neutral in the debate between strong and weak categoricalism, and acknowledge that the
debate could not and should not be settled.

6.4. The argument from modal intuition

I have expressed my view that it is justified to leave the debate between strong and weak
categoricalism unsettled, because I expect the debate to be settled by other researches that take our
best physical theories into account. Nevertheless, an important and influential armchair argument for
strong categoricalism, the argument from modal intuition, is provided by Langton and Lewis, who
are both major authors concerned in this dissertation. Strong categoricalism should be
straightforwardly considered as true, and weak categoricalism should be straightforwardly considered
as false regardless of empirical considerations. To justify my position, there is a need to show that the
argument is unsound.

The argument from modal intuition relies on what Langton (2006, p. 176) calls the Humean
intuition (HI), according to which it is conceivable and thus possible that things have causal
behaviours different from their actual ones. What the argument suggests is that, given that HI is
correct, weak categoricalism must be false. I will argue that AMI is unsound because of a lack of
consideration of possible worlds with two kinds of laws of nature.

6.4.1. Hume’s modal intuition and the argument

The argument from modal intuition (AMI) follows from what Langton (2006, p. 176) calls the
Humean intuition (HI), which is famously pointed out by Hume. According to HI, there is no
necessary connection between things and their causal behaviours, because nothing implies the
‘impossibility of conceiving anything different’ (2003/1748, ch. 1.3.4). Hume provides the example
of ‘a billiard ball moving in a straight line towards another’ (1966/1777, ch. 4.1). There are many
conceivable outcomes other than what physics tells us. The billiard ball may halt and return to where
it started, or it could ‘hop’ over the billiard ball in its path and continue on in a straight trajectory.
AMI argues that, given HI, strong categoricalism must be true:
(1) Things in the actual world can have different causal behaviours in other possible worlds. For example, a sample of common salt does not dissolve in an unsaturated container of water. [From HI.]

(2) If weak categoricalism is true, then it is necessary that things causally behave the way they do in the actual world. For example, a sample of common salt does dissolve in a container of (unsaturated) water.

(3) Weak categoricalism is false. [From (1) and (2).]

For proponents of AMI like Lewis (1986b) and Langton (1998), only strong categoricalism can explain HI. Categorical properties may have different causal behaviours in accordance with different laws of nature because dispositional properties are not intrinsic to them. Consider the intuition that a sample of common salt might not dissolve in an unsaturated container of water in some other possible world. Even though the categorical property to dissolve in an unsaturated solution is instantiated in this possible world, we can imagine the laws of nature being different, such that the common salt is no longer disposed to dissolve. On the contrary, weak categoricalism is considered incompatible with HI. Since dispositions and laws of nature are internal to categorical properties, the same categorical property must causally behave in the same way in every possible world and cannot be otherwise.

A basic formation of AMI like the version I have formulated is provided by Langton. She writes:

[T]hings could be just as they are with respect to their intrinsic properties [i.e. categorical properties], yet different with, respect to their causal powers [i.e. dispositions] – in particular, that if the laws of nature were different, things could have the same intrinsic properties, but different powers. […] These modal intuitions rest on certain assumptions about the contingency of laws of nature, and the contingency of connections – if any – between intrinsic properties and causal powers. These modal intuitions suppose that causal powers do in a sense depend on something other than the way a thing is, in and of itself: that causal powers do depend on something else – but the something else is not simply the existence of some distinct object, but the existence of certain laws. (1998, p. 118, original emphasis)

Langton also invites us to perform a thought experiment with regard to God’s creation. In the scenario, laws of nature are ‘contingent results of divine creation’ (p. 118) because God is free to add, or not add, any new dispositional properties to categorical properties as he pleases. If dispositionalism or
weak categoricalism/dispositionalism are true, then God’s creative act will be constrained (pp. 118-119).

Another version of AMI is provided by Lewis. Rather than inviting us to imagine things having different causal behaviours, Lewis invites us to conceive of (duplicates of) things that are free of their original causal behaviour. He writes:

[M]y objection rests on denying the thesis of essential nomological roles. I support this denial by means of the principle of recombination [‘according to which patching together parts of different possible worlds yields another possible world’ (1986b, pp. 87-88)]. Start with a world where the quark colours and flavours do figure in the laws that are supposed to be essential to them. By patching together duplicates of things from that world, we can presumably describe a world where those laws are broken; yet perfectly natural properties are intrinsic ex officio, and so they never can differ between duplicates. The principle of recombination seems to me very compelling indeed. (1986b, p. 163)

Some readers may worry about what it meant by a ‘thing’ in the consideration of HI and the formulation of AMI. Many would consider a sample of common salt to be an object, not a mere bundle of categorical properties. In fact, Shoemaker, a proponent of strong dispositionalism, responds to AMI by suggesting that it confuses the distinction between an object and a property. Objects, such as a person, can change their properties over time and, similarly, can also have different properties in different possible worlds; a property cannot (1980, p. 119). If we were to change the main character of HI from a thing to a property, the intuition might be much less obvious and widely shared. I acknowledge that this worry is crucial. However, as we have seen, for proponents of AMI like Lewis and Langton, HI is precisely enough to suggest that a duplicate (i.e. an object with all non-relational properties being the same as its counterpart) can have a causal behaviour different from its counterpart’s. Furthermore, both of these authors think that objects are solely constituted of their categorical properties. For the sake of the argument, I remain neutral on the distinction between objects and properties and follow the AMI as it is laid out.

Some readers may also be sceptical about the use of modal intuitions to determine real possibilities (see, for example, Swoyer 1982, p. 210). Some others may simply not possess HI (see, for example, Strawson 2008b, p. 277), or believe that it is a result of ignorance (see, for example, Heil 2003, p. 94). Again, I set aside these issues, as the argument would be a non-starter if we did not
acknowledge HI. I offer up a reply to the argument that does not rely on a distinction between an object and its categorical properties and leaves open questions regarding the reliability of HI.

6.4.2. Why the argument is unsound

My reply targets premise (2). Other than the worlds described by strong categoricalism, strong dispositionalism, and weak categoricalism/dispositionalism, I argue that there are some other kinds of possible worlds. Consideration of these worlds will show that premise (2) is false. Consider the following possible worlds:

**Dual-law worlds:** There are categorical properties with intrinsic dispositional essences and distinct from them a set of laws of nature.

No doubt, these worlds can be described as idiosyncratic. I am only claiming that these worlds are possible, not actual (even though it may turn out to be the case that one of them is indeed the actual world).

In these dual-law worlds, there are three possibilities as to how the two kinds of laws of nature together determine the causal behaviours of things. Among them, two provide possible explanations for why premise (2) is false and appears *prima facie* to be true.

**Possibility One:** The causal behaviour of things is jointly determined by 1) the intrinsic dispositional essences of their categorical properties and 2) a set of laws of nature that is extrinsic to their categorical properties.

For example, a particle X has the intrinsic dispositional essence for a charge of 2eV, while the set of laws of nature governs that all particles X have a charge -1eV. The particle turns out to behave as if it has a charge of 1eV.

**Possibility Two:** The causal behaviour of things is given only by a distinct set of laws of nature. The intrinsic dispositional essences that supervenes on categorical properties are overridden by the extrinsic laws of nature.

For example, a particle X has the intrinsic dispositional essence for a charge of 2eV, while a set of laws of nature governs that all particles X have a charge of -1eV and overrides their dispositional nature. The particle turns out to behave as if it has a charge of -1eV.
**Possibility Three:** The causal behaviour of things is given by the intrinsic dispositional essences of their categorical properties, and this overrides the set of laws of nature that is extrinsic to their categorical properties.

For example, a particle X has the intrinsic dispositional essence for a charge of 2eV, while a set of laws of nature governs that all particles X have a charge of -1eV. The intrinsic dispositional essence of the particle overrides the laws of nature. The particle turns out to behave as if it has a charge of 2eV.

There is no need for these kinds of possibilities laid out here to occur in distinct possible worlds. In fact, they could all co-occur in the same possible world. Individual dispositional properties and the laws of nature interact differently in different things. There is a further possibility, whereby the possibilities laid out here also co-occur with the possibilities laid out with what are described by the original versions of strong and weak categoricalism. With this in mind, there can be many combinations of possibilities that can be classified into many new kinds of worlds (other than the two dual-law worlds and the worlds described by the two main doctrines of categoricalism, and perhaps also by strong dispositionalism), with perhaps many new kinds of doctrines that are required to describe each of these worlds. But the recombination is beside the point. If Possibility One or Two goes through, then the argument from modal intuitions is unsound.

The reason why AMI is unsound should now be clear. Premise (2) is false. Even if the actual world is a strong dispositionalist or weak dispositionalist world, it is possible to still have the correct intuitions that things can behave different than they do actually. The possible world that we conceive by HI may, in fact, be a dual-law world, in which Possibility One and/or Two happen. Things located in the possible world (and perhaps even actually) may have their intrinsic dispositional essences overridden by the laws of nature, or behave in a way that is determined by both their intrinsic dispositional essences and the laws of nature of that world.

**6.4.3. Some further remarks on the argument**

Intuitions vary between people. Not everyone will agree that their modal intuitions can be explained by the dual-law model I provided. One may respond, ‘You explain the modal intuition pointed out by Hume and Langton, but my modal intuition is different and more precise. It has further details that cannot be explained by the dual-law possibilities.’ In fact, Lewis’s modal intuition is one example the dual-law model cannot explain, because he is talking about a world with duplicates of things that are
free of their original causal behaviours. By adding dual-laws into the world, the world is no longer a mere duplication of the things of concern.

I acknowledge that intuition is a subjective matter. Certainly, one can have an intuition with enough details that can resist any alternative explanation. One can even have an intuition that doctrines like strong categoricalism, dualism, or theism, for example, are just true. But it is difficult to see how these intuitions can be appealing. As Stalnaker points out, AMI has little dialectical force against critics of strong categoricalism because the critic has no commitment to whatever the argument assumes (2015, p. 421). Its purpose is, therefore, as I understand it, not to be a knockdown argument against opponents of strong categoricalism, but to help introduce strong categoricalism as a consequence of the metaphysical picture we tacitly assume. The reason to begin with HI but not other intuitions is because AMI requires an intuition that is simple, appealing, and widely accepted enough to begin with. I believe that this understanding of the argument in fact aligns with Langton’s position. As she writes, ‘My impression is that the Humean view is, at present, still closer to being orthodoxy than its alternatives’ (2006, p. 176). Returning to Lewis, one question we might ask whether his modal intuition is as generally appealing as those of the others, such as those of Hume and Langton.

For the sake of argument though, let us assume that Lewis’s modal intuition is appealing. It may be true that, as the objector says, the dual-law possibilities cannot explain Lewis’s modal intuition. Perhaps a more sophisticated consideration of the dual-law possibilities in combination with some independent modal theories can do the job, but it is not our purpose here. What I suggest is that the dual-law possibilities cannot explain Lewis’s modal intuition, but can explain away it. The holder of the intuition might, without awareness, have imagined a set of laws of nature that is extrinsic to their categorical properties. Duplicates of things that are free of their original causal behaviours seem possible because there is a hidden set of laws of nature in the imagined possible world, which has only one role: to suppress those causal behaviours. In this case, we weaken our claim that our approach is neutral to the reliability of HI, but the modal intuition is still explained, though away.

6.5. Conclusion

In response to the question of whether strong or weak categoricalism is true, I suggested that my view on categorical properties implies none of them. Acknowledging the limit of armchair arguments, I expect that that the answer to the question is to be determined by taking our best physical theories into account, which cannot be done in this dissertation. Furthermore, for the purpose of this
dissertation, which is providing an account of the relation between metaphysical naturalism and the
categorical ignorance thesis that can be appealing to most philosophers interested in metaphysical
naturalism, I also purposefully leave the controversies unsettled. In response to Langton and Lewis’s
argument from modal intuition, an influential armchair argument that aims at proving strong
categoricalism regardless of empirical considerations, I have shown that it is unsound.
Part II

Section 7. Categorical ignorance

7.1. Introduction

With our discussions in Sections 4, 5, and 6, we have an idea of what categorical properties are, and an idea of their existence and nature (even though I purposefully left the debate between strong and weak categoricalism unsettled). Taking these into account, we can now discuss the categorical ignorance thesis (CI). The reader is reminded that CI is the following:

CI: Categorical properties are in principle unknowable. All of our knowledge is, in fact, about causal, dispositional, and locational properties.

We have seen in Sections 5 and 6 that many philosophers – categoricalists and strong dispositionalists included – believe that categoricalism implies CI. Categoricalists who believe in CI include Rae Langton (1998), David Lewis (2009), Frank Jackson (1998), Simon Blackburn (1990), and many others. Strong dispositionalists who believe that categoricalism implies CI include Sydney Shoemaker (1980), Chris Swoyer (1982), Brian Ellis and Caroline Lierse (1994), Robert Black (2000), John Hawthorne (2001), Alexander Bird (2005), and many others.

In this section, I will assess the major arguments for CI. While I will defend the doctrine, I will argue that, contrary to the thought of some, no simple line of armchair philosophical arguments can show that it is true in any compelling way.

The structure of this section will be as follows. In Section 7.2, I will offer an understanding of CI that I find to be the most tenable. In Section 7.3, I will defend CI, providing my argument for it. In Sections 7.4, 7.5, and 7.6, I will critically assess other arguments for CI in the literature, which are offered by Langton, Lewis, and Jackson.

7.2. The nature of CI

To discuss whether CI is true, we must first determine what the thesis really is. For some philosophers like Sam Cowling, CI is the view that we cannot know of any fact about categorical properties (2010, p. 662). But this version of CI is obviously – rather too obviously – implausible. On the one hand, as
Cowling argues, this is self-defeating because we cannot even know that categorical properties are properties and that they fall prey to CI (p. 662). On the other hand, it is difficult to see why, as long as we can assume that categorical properties could exist, we cannot ascribe necessary attributes such as being self-identical to them, since these necessary attributes are supposed to applied to every logically or metaphysically things without exception. In addition, we can also probably speculate that some contingent attributes such as ‘not being Mickey Mouse’ can be ascribed to categorical properties.

One more interesting kind of possible knowledge of categorical property is worthy of mentioning before we go on. Ann Whittle argues that we can identify a particular categorical property \( x \) as the realiser of a dispositional role \( y \) in the actual world, much like we can identify the person Napoleon solely by the role he plays in history (2006, pp. 469-470). Taking this into account, Whittle argues that we can know of the categorical property \( x \), much like we can know of the person Napoleon. This kind of knowledge of categorical properties, which Lewis calls reference-fixing (2009, p. 216), cannot easily be refuted.

It is easy to reject CI if it is assumed to be the implausibly strong doctrine that we cannot justifiably and correctly ascribe any attribute, or any of the above attributes, to categorical properties. But this is not a charitable way to assess the doctrine, because what counts as knowledge is highly context-sensitive.\(^{29}\) Hence, the more charitable way to assess CI is to make it clear what kind of knowledge CI is really targeting before we judge whether it is correct. As Tom McClelland (2012) replies to Cowling and as Dustin Locke (2009) replies to Whittle, the kinds of knowledge they discuss are not really what CI concerns. The same can presumably be said about the other kinds of knowledge of categorical properties I listed above.

What then is the exact kind of knowledge CI is targeting? The answer to this question varies among theorists of CI because they all have different views on the nature of categorical properties, the conception of knowledge, the way CI consists, and so on. I shall not assess their philosophical systems one by one at the moment. This is because I have already provided my own views on the nature of categorical properties and thus should focus on it. In addition, as I have mentioned earlier, in order to allow my investigation of the relation between metaphysical naturalism and CI to be appealing to more metaphysical naturalists, controversial assumptions that are unnecessary should be avoided. With this in mind, I will straightforwardly develop my own version of CI, with the aim that

---

\(^{29}\) Dustin Locke, a proponent of CI, provides a detailed discussion of the issue (2009, p. 228).
its only assumptions are my view on the nature of categorical properties and some philosophical assumptions that are appealing to most philosophers.

We return to the question of exactly what kind of knowledge CI is targeting. To answer this question, we have to recall my views on the nature of categorical properties. In Sections 5 and 6, I developed a version of the thick quiddity picture in the light of a range of considerations of metaphysics and natural science. The thick quiddity picture, as David Chalmers defines it, is the view that categorical properties do have substantive natures or features and are not, as the thin quiddity picture describes, featureless. Taking this into account, CI, in my view, is that we cannot know of such substantive natures or features, in the sense that we cannot know what they are precisely like.

Let me illustrate what is meant by an example. In the case of a physical structure of a plastic bottle, we can tell some of its very precise and concrete details, such as the curvature of its surface, the molecular structure of a polyethylene molecule, and so on. These details are different from the abstract categories that the plastic bottle falls within (such as being self-identical) and from the details about the possible influences of the plastic bottle (such as what it can cause or realise). Let us return to the case of categorical properties. It seems that this kind of detail is ungraspable. Even though we can, of course, know of them as the player of such and such roles in science, the realiser of such and such things we know of, a member of the metaphysical category ‘being self-identical’, and so forth, none of these are really describing what they are precisely like in the way I mentioned above.

As I have mentioned, some theorists of CI might have other views on the nature of categorical properties. For example, according to Chalmers’s interpretation of Lewis, Lewis’s view is one that is ‘very close’ to a thin quiddity picture, according to which categorical properties are featureless (2012, p. 349n19). If Chalmers is correct, Lewis’s version of CI should target different kind of knowledge of categorical properties. This is because there is no possible knowledge of the substantive natures or features of categorical properties, except the knowledge that there are no such natures or features, which Lewis does know of. I remain neutral on how correct Chalmers’s interpretation of Lewis above is, because interpreting Lewis’s works is sometimes not an easy job. To be more precise, many of Lewis works are best understood in relation to his other works, and those relations are often open to interpretation. Much like the case of Lewis, because of our purpose and for the limit of scope, it is not possible to analyse the works of each theorist of CI in order to determine the correct interpretations of them. With this in mind, when I discuss the arguments for CI of other authors, I will assume that

---

30 It is noteworthy that these details are graspable as knowledge in multiple ways. Sometimes we can express it as sentences, and sometimes we can only have mental representations of them in our mind. I shall not discuss this further.
they are for my version of CI and assess whether or not they confirm it; and their own views on the nature of categorical properties and their own versions of CI will be set aside.

7.3. My argument for CI

I have offered my version of CI in the previous subsection. In this subsection, I will explain my motivation to believe in it. My motivation to believe in CI is not any simple line of armchair philosophical argument, but is a result of a combination of many hypotheses about what the world is like, including those in metaphysics, science, and epistemology. I call it the combination of hypotheses argument.

Let me illustrate. Ultimately, whether we can know of a kind of property depends on what (1) we ourselves, (2) the kind of property of concern, (3) the ontological relations between them, and (4) the epistemic relations between them are, or could be, like. We know of the nature of these things by making numerous hypotheses in metaphysics, natural science, and epistemology, such as our hypotheses about what a human subject is like, what our sensory system is like, what the physical world is like, what a property is, what counts as knowledge, and so forth. Only when we take a substantial number of these hypotheses into account, can we be sufficiently informed to make a judgment concerning whether human knowledge can in principle grasp a kind of property, and thereby confirming or disconfirming an irremediable ignorance thesis like CI.

Returning to the issue of CI, I believe that the thesis is true. But unlike some proponents of CI, my view is not derived from any simple line of armchair philosophical argument. Instead, it is simply based on my observation that when we combine our best hypotheses in the many areas mentioned above, we cannot derive any tenable hypothesis that can allow us to know of categorical properties. No doubt, this observation could be mistaken, since there could be ways to develop a tenable combination of tenable hypotheses that can allow us to know of categorical properties. Here I list three possibilities why it could be mistaken. The first possibility is that it simply has not been discovered by any philosopher (or scientist); the second possibility is that it has been discovered, but I am not yet informed; the third possibility is that it has been discovered and I am informed, but it is based on some hypotheses that many of us mistakenly take to be untenable, such as, say, a version of Hindu idealism. Nevertheless, absence of evidence in an all-round consideration can be evidence of absence, and the all-round consideration mentioned here is not done only by me, but also many other metaphysicians in the area. This is shown by the fact that while CI is undergoing intense research, (almost) none have mentioned ways to know of categorical properties (with the exception of knowing
of them through our qualia, which I will discuss in detail in Part III). Of course, in line with what has been argued, I must open to any future research that could provide a hypothesis that shows otherwise and assess each carefully. This is because it is difficult to deny the possibility that opponents of CI will eventually be able to provide a new combination of hypotheses that can allow for knowledge of categorical properties. In particular, the best hypotheses in some important areas such as physics are often radically modified or replaced. Nevertheless, this openness to revision is not in conflict with my current conclusion that CI is true, just as an Einsteinian physicist’s openness to revision of physical theories is not in conflict with her belief that nothing in principle can move faster than the speed of light.

A matter of concern is then whether the combination of hypotheses argument can really be derived from my set of hypotheses about what the world is like. In Sections 5 and 6, I developed a version of the thick quiddity picture in light of a range of considerations of metaphysics and natural science. In this view, a (fundamental) categorical property does have features or natures that realise the non-fundamental categorical properties of a fundamental object like mass, charge, size, and so forth.

For many, the picture might seem to imply that we have knowledge of categorical properties, such as knowledge of its mass, charge, size, and so forth. But this does not seem to be true. Before the discussion begins, it is noteworthy that in Section 5.5.2 I argued that properties like mass, charge, size, and so forth cannot be fundamental but are, instead, non-fundamental by the argument from correlation. With this in mind, in the view that I actually defended, the precise relation between the non-fundamental categorical properties and the features or natures of the categorical property that realise them is a realisation relation. Taking this into account, knowledge of mass, charge, size, and so forth does not necessarily imply knowledge of categorical properties, because knowledge of the realised does not necessarily imply knowledge of the realisers. No doubt, a possibility is that we may

31 Johnathan Schaffer has an article called ‘Quiddistic knowledge’ (2005). Unfortunately, the article is simply a reply to Lewis (2009), attempting to show that Lewis’s argument is not compelling. In the article, Schaffer argues that Lewis’s argument from multiple realisability for CI (which I shall discuss in Section 7.6) is a version of the arguments for scepticism about the external world, and has no novel structure of argument. Hence, traditional replies to scepticism about the external world can apply to the argument. What I believe he has not provided is an exact example of knowledge of categorical properties, or a way to acquire such knowledge.

32 A possible exception is the view that spatiotemporal properties are a kind of categorical property (see Ellis and Lierse 1994; Ellis 2010). It might seem that we can know of the natures or features of spatial properties and temporal properties. I shall set issues of spatiotemporality aside in this dissertation because they require an independent investigation.

33 The reader is reminded that the categorical properties I have been mentioning along the way are assumed to be fundamental unless specified, since they are the kind of categorical property this dissertation is interested in (see Section 4.4).
speculate about what a realiser is like by using our knowledge of the realised as the starting point of inference to the best explanation. However, it is difficult to see how inference to the best explanation can work in our case. In current physics, non-fundamental categorical properties like mass and size are now nothing more than sets of very abstract numerals in some very abstract theoretical models. It is true that they could reflect the quantities of some features or natures of categorical properties that bear them. But the problem is that it is difficult to speculate about exactly what kind of features or natures have those quantities. In other words, what we know at most is that an unknown *quality* is having such and such *quantity*. It might be argued that we do have the knowledge of the features or natures of categorical properties, because we can know that, say, ‘the substantive feature x of a categorical property y has a quantity of eight’. I do not wish to deny this kind of knowledge, but I would note that it is terribly imprecise, and that CI still applies to any more precise knowledge of categorical properties. In this sense, it could be said that I commit to a weaker version of CI.

Presumably, not every philosopher agrees that CI has to be motivated by a complex all-round consideration like my combination of hypotheses argument. Some believe that some simple armchair arguments can show that CI is true. For example, Langton, Lewis, and Jackson offered the *argument from irreducibility* and the *argument from multiple realisability*. Details of these arguments will be explained as the discussion goes on. In the rest of this section, I will argue that the arguments are unsound, unattractive, or question-begging. The reason behind this, I believe, is that many phenomena in the world have to be explained by a combination of many factors or laws, not any particular single factor or law; and if CI is such a phenomenon, then simple armchair arguments may not be able to take the amount of factors or laws into account.

It is noteworthy that my view is not that all simple armchair arguments are useless. On the one hand, some of them may be used as useful examples to guide people to be aware of the fact that their views imply CI. This is important because some people might have already tacitly committed to a combination of hypotheses that implies CI, but are not aware of the fact. Simple armchair arguments, in this case, can help these people to acknowledge the implications of their own beliefs. On the other hand, with a set of hypotheses about what the world is like, those arguments might point out some hypotheses that play a more crucial role in encompassing CI than most of the others. My point is simply that such explanations can hardly be full or crucial explanations because of the multiplicity and complexity of the factors involved.
7.4. Langton’s argument from irreducibility for CI

7.4.1. The argument

Langton’s argument from irreducibility for CI begins with a description of our sensations. In her view, we know about external things only though ‘receptivity’, in which external things affect us and thus allow us to form representations of them. She writes:

**Receptivity:** Human knowledge depends on sensibility, and sensibility is receptive: we can have knowledge of an object only in so far as it affects us. (Langton 1998, p. 125)

With the receptivity thesis, Langton argues that we can in principle have knowledge of whatever affects us (p. 126).

The next step of the argument is to point out what actually can affect us – something that can only happen causally. In her view, things causally affect each other only by their causal powers. Thus, what we can in principle know about things is only their causal powers, which are, in turn, commonly described as the dispositions of things in contemporary analytic metaphysics. As she notes, she uses the terms ‘power’ and ‘disposition’ synonymously (p. 110). Taking this into account, we can only know of the dispositions of things. I call this view epistemological dispositionalism. With the above in mind, the first half of the argument from irreducibility can be stated as the following (note that the argument is unsound):

(1) We can have knowledge of an object only in so far as it (causally) affects us. (Receptivity.)
(2) Things (causally) affect each other only by their causal powers. (Premise.)
(3) Therefore, one can only have knowledge of causal powers. (From [1] and [2].)

In fact, (3) does not follow from (1) and (2). But I will set it aside at this moment and return to it when I assess the whole argument. Now I shall introduce the second half of the argument.

Langton believes that causal relations must have bearers. Langton cites Kant, ‘Concepts of relation presuppose things which are absolutely [i.e. independently] given, and without these are impossible’ (p. 16). In her view, the bearers have to be what she calls intrinsic properties, which, as we have seen in Section 5, are what are standardly called categorical properties in contemporary metaphysics. However, she also believes that the categorical properties play no role in causal relations other than acting as their bearers. This is based on her irreducibility thesis:
Irreducibility: The [causal powers] of substances are not reducible to the [categorical properties] of substances. (p. 109)

In her view, the laws of nature (for example, those governing Newtonian attraction) are not necessary consequences of categorical properties. After God has created categorical properties, the laws of nature do not follow but are the contingent result of a further divine creation (p. 118). Hence, the irreducibility thesis is true.

It is not difficult to see that the irreducibility thesis is actually strong categoricalism (which we have discussed in Sections 6 and 7) in disguise. In fact, Langton does make it clear that the two theses are closely related (p. 120n32). The reader is reminded that strong categoricalism is the following:

**Strong categoricalism:** All fundamental properties are categorical properties without dispositional essences intrinsic to them. For there to be dispositional properties, categorical properties are required to act as their bases. But the relation between the two kinds of properties (i.e., which categorical property is playing which dispositional role) is determined by the laws of nature – or what described by the laws if the laws are understood to be descriptions – that are extrinsic to categorical properties, not the intrinsic essences of categorical properties.

Langton’s argument for strong categoricalism is the *argument from modal intuition*, which I have already discussed and criticised in Section 6. I shall not repeat the introduction to and assessment of the argument here.

With the above in mind, the argument from irreducibility in its completed form can be put as follows:

(1) One can only have knowledge though one’s being (causally) affected. (Receptivity.)
(2) Things (causally) affect each other only by their causal powers. (Premise.)
(3) Therefore, one can only have knowledge of causal powers. (From [1] and [2].)
(4) Causal powers are not reducible to categorical properties. (Irreducibility.)
(5) Therefore, categorical properties are unknowable. (From [3] and [4].)

Putting everything together, the idea is that if causal powers are not reducible to the categorical properties, they are then distinct things. However, according to sub-conclusion (3), i.e.
epistemological dispositionalism, we can only know of causal powers. It then follows that epistemological dispositionalism rejects any knowledge of categorical properties.

It is noteworthy that the move from sub-conclusion (3) and premise (4) to conclusion (5) is not valid. This is because what Langton does not consider here is a reduction in the opposite direction, which is a reduction of categorical properties to causal powers. However, the lack of consideration of the scenario is understandable, because the kind of categorical properties of concern here, as I have argued in Section 4, are properties that are categorical, intrinsic, and fundamental. Therefore, categorical properties are by definition irreducible to causal powers. The scenario Langton misses is simply ignorable.

The overall structure of the argument from irreducibility is demonstrated by Figure 1.

![Figure 1. Langton’s argument from irreducibility](image)

7.4.2. Critical assessment of the argument

In this subsection, I will discuss two steps of the arguments that I believe to be controversial or mistaken.
7.4.2.1. Premise (4): the irreducibility thesis, or strong categoricalism

The irreducibility thesis, or strong categoricalism (as most contemporary metaphysicians call it), is a thesis that many categoricalists do not accept. We have seen in Sections 6 and 7 that many philosophers are weak categoricalists who believe that dispositions are reducible to their categorical bases. I have also discussed Langton’s argument from modal intuition for strong categoricalism in Section 6, and have argued that it is unsound. I shall not go further into this issue because it has already been discussed in detail.

7.4.2.2. Sub-conclusion (3): epistemological dispositionalism

I consider it to be the most dubious move in the argument that sub-conclusion (3) follows from premises (1) and (2). Recall that premise (1) is the receptivity thesis, according to which one can only have knowledge though one’s being (causally) affected; premise (2) is that things (causally) affect each other only though their dispositions; and conclusion (3) is epistemological dispositionalism, according to which one can only have knowledge of causal powers/dispositions. The problem is that sub-conclusion (3) simply does not follow from premises (1) and (2). To make sense of the move, what we need is not actually premises (1), i.e. the receptivity thesis, but another premise, ‘(1’) we can only know of what (causally) affects us’.

Premise (1’) is not implied by the receptivity thesis but should be viewed as an independent thesis. In her book, Langton has claimed that the receptivity thesis implies that we can in principle know what affects us (1998, p. 126). This might be true, but this, too, does not imply premise (1’). The receptivity thesis merely tells us that we need causation as a starting point to know about the world, and it does not tell us that it is the end point.

My view is that premise (1’) is obviously false, and that the receptivity thesis and premise (1’) must not be confused. Like the friends of the razor arguments I criticised in Section 5.5.3, Langton seems to have assumed an implausibly narrow conception of empirical explanation, which accepts only the kind of causal explanation that traces causal connections and then explains an effect by its cause(s). But there are also structural explanations in science, which tell us why the causal connections in the world are arranged in the way we perceive them. When we provide structural explanations we do not need to be causally connected to the explanans. The method required here is simply inference to the best explanation, whereby we choose the most likely possibility by a full range of theoretical virtues, such as simplicity, goodness to fit with data, predictivity, and so on. For
example, we have seen in Section 5 that science does provide binding explanations, which are (1) the explanations of colocations of multiple dispositions through positing their common bearer objects, or (2) the explanations of correlations of multiple dispositions through positing some natures of their common bearer objects, such as size, mass, charge, and so forth. In sum, premise (1’) is simply in conflict with our practice of science.

Besides, the fact that we can know about non-fundamental categorical properties like size, mass, charge, and so forth also tells us that sub-conclusion (3), i.e. epistemological dispositionalism, cannot be true. I believe that it can be concluded that epistemological dispositionalism seems to be implausible.

With the above in mind, the argument from irreducibility is unsound. Langton makes use of the irreducibility thesis to show that knowledge of categorical properties is incompatible with her criterion of empirical knowledge, which is epistemological dispositionalism. But I have shown that epistemological dispositionalism is implausible, let alone the controversy surrounding her irreducibility thesis (or strong categoricalism).

7.4.3. Some additional remarks on the argument

We have seen that Langton’s argument from irreducibility is unsound. It is not easy to see how the argument can be fixed. To be fair though, Langton tells us the important fact that our ability to know about the empirical world is limited by receptivity. With regard to her basic position that there might be some features of things we cannot know of because of the epistemic limit that receptivity provides us, she might, indeed, be correct.

However, if we sympathetically understand Langton in this way, it seems to me that we are motivated to return to the combination of hypotheses argument. We have seen that the receptivity thesis cannot combine with epistemological dispositionalism to formulate a simple armchair argument for CI. What else can we do with the receptivity thesis? One way is to formulate another simple armchair argument by the thesis, but I have not discovered any such argument. In my view, the approach with a better chance of success is to understand it as a component of our framework of epistemology, which is, in turn, a component of our understanding of the world. It is our understanding of the world that tells us whether we can know of categorical properties, and the receptivity thesis is only a small part of it. This approach is exactly in line with how the combination of hypotheses argument operates.
7.5. Lewis’s argument from the semantics of science

In his seminal article ‘Ramseyan humility’ (2009), Lewis offers an argument for CI, which, unlike Langton’s argument, is based on an analysis of the semantic structure of scientific theories. The argument relies on (1) the Ramsey-Lewis method of defining theoretical terms and (2) multiple realisability (which I will explain). His argument is that (a) (1) implies (2), and that (b) (2) in turn implies CI. I will assess (b) separately in Section 7.6, and focus on (a) in this subsection. With this in mind, let us assume that (b) is true. In this subsection, I will focus on (1) and (a). I will argue that Lewis does not reach his conclusion solely from the Ramsey-Lewis method, but also a number of metaphysical assumptions that are significantly controversial and are facing strong empirical challenges.

Lewis begins with his model of the semantic structure of scientific theories, which he develops in his early and seminal article ‘How to define theoretical terms’ (1970). He is a scientific realist. He asks us to assume that there is a final scientific theory T about the natural world. In his view, theory T, like all other scientific theories, consists of O-terms and T-terms. O-terms are the terms that are used in our older and ordinary language outside theory T; T-terms are theoretical terms that are specifically defined in theory T. Each T-term has to be defined holistically in relation with other T-terms by O-terms. Such relations include their nomological and locational roles in theory T (p. 207). Some of the nomological roles would be played by categorical properties, while he assumes that none of these properties will be named in O-terms. He writes, ‘The fundamental [categorical] properties mentioned in T will be named by T-terms. I assume that no fundamental [categorical] properties are named in O-language, except as occupants of roles’ (2009, p. 206). Although Lewis does not make it clear why he assumes so, one of his other works (1972) seems to hint that he supposes that the use of O-terms is to name and define nomological and locational relations. With the assumption that the roles played by categorical properties are identified solely in relational means, Lewis argues that they are multiply realisable by multiple categorical properties. He then takes his last step, ‘If there are multiple possible realizations, [CI] follows’ (2009, p. 207). (Recall that it is our current assumption that multiple realisability implies CI.)

---

34 As I have noted in n3, there are multiple interpretations of Lewis’s article because there are different ways to read the article in relation to Lewis’s other works. Chalmers’s interpretation is that Lewis is committed to a thin quiddity picture, according to which a categorical property is featureless, with the exception of its numerical identity. Hence, Lewis’s version of CI is an ignorance of numerical identity, not qualitative ignorance (Chalmers 2012, p. 349n19; see also Stoljar 2015a). This dissertation is not one on Lewisian scholarship. With this in mind, as I have mentioned earlier, I shall not investigate which interpretation of Lewis is correct. I will simply assume that his argument is one for qualitative ignorance, which is set as the target of investigation of this section, and determine whether the argument is successful or not.
In my view, it appears that what Lewis aims to do is analyse a theory of the semantics of science, which he has developed at an abstract level from the armchair, and investigate its metaphysical implications. However, if this is what Lewis really aims to do, then it is not obvious that his analysis perfectly matches his purpose, at least if he has the intention to develop an analysis that is compelling to most, not only to the few who share all of his other views. This is because a number of metaphysical assumptions about what the world is like are, in fact, made within his investigation. In particular, he seems to have assumed that his doctrine of ‘Humean supervenience’ is true. He introduces the doctrine as follows in an earlier article (1986b):

It is the doctrine that all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another. (But it is no part of the thesis that these local matters are mental.) We have geometry: a system of external relations of spatiotemporal distance between points. Maybe points of spacetime itself, maybe point-sized bits of matter or aether or fields, maybe both. And at those points we have local qualities: perfectly natural intrinsic properties which need nothing bigger than a point at which to be instantiated. For short: we have an arrangement of qualities. And that is all. There is no difference without difference in the arrangement of qualities. All else supervenes on that. (pp. ix-x)

It is noteworthy that Lewis, in that article, has famously noted that the doctrine is the assumption of most of his works (p. ix). But this alone, of course, does not show that the article ‘Ramseyan humility’ has also made the assumption. My belief that Lewis has made the assumption is based on two reasons. The first is that we have seen that Lewis assumes that no description of nomological and locational relation in theory T is played by a non-relational categorical property. It is not easy to see why, say, a locational relation such as a length of 15 cm in theory T cannot be played by a non-relational categorical property. It is not easy to see why, say, a locational relation such as a length of 15 cm in theory T cannot be played by a non-relational categorical property. In fact, similar views have been defended by J. J. C. Smart (1963) and Alyssa Ney (2015). For example, Smart writes:

[W]e could explore the possibility of giving a theory of length, mass, and so on, as absolute and not relational. […] We do indeed test propositions about length relationally, but that to go on to say that length is purely relational is to be unduly verificationist about meaning. (1963, p. 74)
Of course, whether the kind of scenario Smart describes is logically and metaphysically possible, let alone whether there are actually categorical properties like those within, is worthy of an independent inquiry. But it is difficult to see why it should be left ignored. One possible explanation is that Lewis assumes his doctrine of Humean supervenience, according to which categorical properties are absolutely ‘little things’ that occupy no extended space or time. With this in mind, the possibility that the role of a length of 15 cm in theory T is played by a categorical property is simply ignored.

Setting the above worry aside, there is another important reason why I believe that Lewis assumes that the doctrine of Humean supervenience is true. The reason is that there might be irreducible relational properties if Humean supervenience is not true. This is inconsistent with Lewis’s Ramseyan argument which appeals to multiple realisability, because some kind of theoretical role can be exhausted by a single irreducible relational property. For example, imagine a theoretical term A that means ‘a fundamental property that is solely an equilateral triangle with size x’. Any property that plays its role, according to the Lewisian framework, must be an irreducible relational property. In addition, because of the term ‘solely’ in the description, any two fundamental properties that can play the role of term A must have exactly the same nature, and hence the role of term A is not multiply realisable.

It might be suggested that many of the above assumptions can be assumed even in an analysis of a model of the semantics of science at the abstract level because they are logically or metaphysically necessary. For example, many philosophers consider fundamental relational properties to be metaphysically impossible. One common supporting argument is that relations require relata (see, for example, Jackson 1998, p. 24). Others consider fundamental geometrical properties to be metaphysically impossible because all geometrical properties can be possessed by empty space. In their view, there must be a further property to ‘fill in’ the space, or else there will be nothing that distinguishes the geometrical property from an empty space (see, for example, Armstrong 1960, pp. 185-187; Langton 1988, pp.165-166).

However, in my view, these arguments for the metaphysical impossibility of fundamental geometrical properties, like the doctrine of Humean supervenience, are philosophical concerns independent of the semantics of science. Nothing in the semantics of science implies them. Hence, they are metaphysical issues that should be assessed elsewhere. Moreover, a much more important concern is that it is not Lewis’s view that the properties mentioned above are impossible. Unlike many philosophers above, he, in fact, considers fundamental relational properties to be possible. With
this in mind, he also considers the question of whether or not Humean supervenience is true to be an empirical issue. He writes:

For I concede that Humean supervenience is at best a contingent truth. Two worlds might indeed differ only in unHumean ways, if one or both of them is a world where Humean supervenience fails. Perhaps there might be extra, irreducible external relations, besides the spatiotemporal ones. […] It is not, alas, unintelligible that there might be suchlike rubbish. Some worlds have it. (1986b, p. x)

I have conceded that Humean supervenience is a contingent, therefore an empirical, issue. (p. xi)

This means that Lewis’s semantics of science leads to CI simply because the empirical physical world contingently turns out to be like such and such. The worry here is as follows: Lewis’s analysis is based on an armchair model of the semantics of science that is at the abstract level, but not on a set of semantics of science that has been historically developed in relation to concrete empirical facts. With this in mind, his analysis should not assume too much about what empirical facts will turn out to be like, unless it has very reliable support from contemporary scientific theories. Unfortunately, there is a growing number of philosophers who consider the doctrine of Humean supervenience to be false because of contemporary scientific discoveries (see, for example, Oppy 2000; Esfeld 2004; Ladyman & Ross 2007). Even Lewis himself admits that he is aware that there is a challenge from physics to the doctrine of Humean supervenience and is ready to give it up. What he demands is simply that the relevant physical theories have to be shown to be not merely instrumentalist (1986b, p. xi). I consider that it is, indeed, difficult to see why anyone should commit to the doctrine of Humean supervenience if she takes science to be more reliable than mere armchair speculations. The worry is more significant when Lewis asks us to make suppositions of the final scientific theory. It seems that there can be hardly any reason to suppose that the final scientific theory will align with the doctrine of Humean supervenience.

To conclude, we have seen that Lewis’s model of the semantics of science does not, by itself, entail the truth of the multiple realisability thesis. When Lewis defends CI, he also makes use of a number of metaphysical assumptions that are both controversial and which face strong empirical challenges. Hence, it is not the case that his model of the semantics of science plays the central role in his argument. In my view, what Lewis has actually given us is an attempt to combine several
hypotheses in different areas into a single hypothesis, in order to see if knowledge of categorical properties is possible. These hypotheses include strong ones on what the world is like, rather than exclusively hypotheses concerning our cognitive capabilities or the scientific method. The project makes sense for Lewis’s systematic philosophy: it is indeed plausible that knowledge of categorical properties is impossible in the complete Lewisian worldview. But a worry for a more general project of determining whether there can be knowledge of categorical properties is that, of course, some hypotheses Lewis uses are very strong assumptions and are not obviously attractive; and even if they are, there might not be many who share exactly the same set of hypotheses. A more significant worry is the number of factors that are actually required in determining whether CI is true, especially if we agree with philosophers like James Ladyman and Don Ross (see Section 5.7) who argue that metaphysics should be based on physics, not an armchair approach like Lewis’s. (We should remember that, for example, there are more than ten interpretations of quantum physics in theoretical physics, and it is currently difficult to see which one is true or best.) This complexity makes it *prima facie* difficult to determine whether CI is true on the basis of an approach like Lewis’s.

On the other hand, my combination of hypotheses argument does not require us to begin with any particular strong thesis in metaphysics (or natural science). Rather, it invites us to play the role of an observer, placing the burden of proof on those who claim that there could be methods of acquiring knowledge of categorical properties. Hence, we simply need to wait for research projects that attempt to discover such methods, and assess them when they are proposed. The fact is that, with the possible exception of knowing of categorical properties through qualia (which I will discuss in Part III), no plausible example of knowledge of categorical properties has been provided by any combined hypotheses. In this sense, since my combination of hypotheses argument does not share the narrowness of Lewis’s approach, it should be both more convincing, and also appeal to a broader range of philosophers.

### 7.6. The arguments from multiple realisability

In Section 7.5, I showed that Lewis’s account for the relation between categorical properties and the semantics of science has a considerable limitation. But the limitation of Lewis’s account does not show that the arguments from multiple realisability (offered by him and others) face the same limitation, because they do not necessarily need to take Lewis’s view on the semantics of science as an assumption. For example, Jackson (1998) does not provide an analysis of the semantics of science but nevertheless offers an argument from multiple realisability, which I shall discuss. What the
advocates of the arguments need to show is only as follows. Some posits of our science are roles that require realisers, and there could be more than one possible realiser that can match the descriptions of the roles. There is then multiple realisability. Since the science cannot distinguish these possibilities, this follows that it cannot contain knowledge of such properties.

In the following part, I will introduce three versions of the arguments from multiple realisability, which include (1) the permutation argument, (2) the replacement argument, and (3) the indistinguishability argument. I will then provide an overall assessment of the arguments from multiple realisability. More precisely, I will argue that they are all question-begging and unnecessary but may be useful in convincing people to believe in CI.

7.6.1. The permutation argument

The permutation argument is offered by Lewis. It begins with two basic assumptions: combinatorialism and quidditism, which Lewis very briefly states in the article ‘Ramseyan humility’:

Combinatorialism tells us that the laws of nature are contingent. (2009, p. 209)

Quidditism is the premise that tells us that the permutation is […] a different possibility. (2009, p. 209)

Lewis’s statement seems to be expressing the implications of the two doctrines rather than what they actually mean. What the two doctrines actually mean, as I understand them, is as follows. Combinatorialism follows from Lewis’s principle of recombination (which he discusses elsewhere), according to which ‘patching together parts of different possible worlds yields another possible world’ (1986a, pp. 87-88). Lewis argues that the yielded possible world is presumably one ‘where those laws are broken’ (1986a, p. 163). For this to happen, the laws of nature must be contingent, not something that is necessarily attached to categorical properties.

Quidditism is the doctrine that nomological roles have categorical properties as their realisers. With this in mind, a possible world with the same set of nomological roles and the same set of categorical properties as the actual world may be a world that is not identical to it, because a categorical property may take up a different nomological role.

From the above two assumptions, Lewis argues that a scenario in which the players of two actual nomological roles are swapped will not change anything else, including the nomological roles they play and the locations they occupy. Therefore, a swapping of realisers is merely another possible
realisation of our scientific theory. Acknowledging that our science cannot distinguish between the actual realisation of nomological roles by their categorical bases and its permutations, it can be concluded that we are in principle ignorant of categorical properties.

7.6.2. The replacement argument

The replacement argument, unlike the permutation argument, does not appeal to an exchange of roles. Instead, the argument begins with the assumption that players of the T-term roles are replaceable by what Lewis calls *idlers* and *aliens*. There is then multiple realisability. Idlers, according to Lewis, are fundamental properties that are within the actual world, but play no nomological role. Aliens are fundamental properties that only exist in unactual possible worlds (2009, p. 205).

Again, Lewis argues that a replacement of the players of nomological roles by idlers and aliens will not change anything else. It is another possible realisation of our scientific theory. Acknowledging that our science cannot distinguish between the actual realisation of nomological roles by their categorical bases and its replacements, it can be concluded that we are in principle ignorant of categorical properties.

7.6.3. The indistinguishability argument

‘The indistinguishability argument’ is a title I attribute to an argument offered by Jackson (1998). In support of what he calls ‘Kantian physicalism’, which is actually CI, Jackson offers the indistinguishability argument, which is an argument that is very similar to Lewis’s two arguments. The argument is as follows. There might be two categorical properties that are not identical, but they realise the same nomological role in our science in succession (1998, pp. 23-24). A conclusion much like that of Lewis then follows: acknowledging that we cannot determine whether a nomological role is played by one or two categorical bases, it can be concluded that we are in principle ignorant of categorical properties.

7.6.4. An assessment of the arguments

I believe that the most significant problem of all versions of the arguments from multiple realisability is that they are question-begging. The reason is as follows. For theoretical roles in our science to be multiply realisable, our science must more or less lack some constraints on what the realisers of the theoretical roles are. If our science can contain perfect and precise information on what the realisers of the theoretical roles are (and perhaps, in addition, methods to verify the information), the
information should act as a constraint that strictly restricts the theoretical roles to be played by their actual realisers. Hence, when we judge that there is a multiple realisability of our science as the arguments suggest, we must have already (tacitly or explicitly) judged that some information of the realisers must be missing in our science. In sum, to judge that multiple realisability is true, we must have already judged that CI is true. This is question-begging.

Note that the fact that the arguments from multiple realisability are question-begging does not imply that they are useless. This is because some people might have implicitly judged that CI is true without being aware of it. In this case, when they discover that they could accept multiple realisability by categorical properties, they should also discover the fact that they have already implicitly judged that CI is true. In sum, the arguments from multiple realisability may help some people to acknowledge their own implicit belief in CI.

7.7. Conclusion

In this section, I defended CI by my combination of hypotheses argument, which appeals to the actual fact that no combinations of hypotheses are yet able to offer us interesting claims of knowledge of categorical properties. I have also assessed some simple lines of armchair philosophical arguments for CI and argued that they are problematic.

As I have mentioned in this section, the fact that there are no interesting claims of knowledge of categorical properties has an exception. The exception is the hypothesis that it is possible to know of categorical properties through our qualia. I will assess it in Part III.
This page intentionally left blank
Part III

Russellian Monism
This page intentionally left blank
Part III

Section 8. Russell and Stoljar’s Russellian Monism

8.1. Introduction

The purpose of this dissertation is to inquire into the connection between metaphysical naturalism (MN) and the categorical ignorance thesis (CI) and determine whether the latter will challenge the former. In Part I, I spelled out what I believe MN to be. In Part II, I discussed the nature of categorical properties and argued that CI is true. This allows us to inquire into the connection between the two doctrines in Part IV, which is the final part of this dissertation.

Nevertheless, an influential theory in philosophy of mind, Russellian Monism, is worthy of a lengthy assessment before we move to Part IV because of its discussion of categorical properties. The doctrine consists of three doctrines, which may be considered independent. They are what I call (1) the scientific categorical ignorance thesis (SCI), (2) the categorical qualia thesis (CQ), and (3) the common bases thesis (CB):

Russellian Monism: (1) The scientific categorical ignorance thesis (SCI), (2) the categorical qualia thesis (CQ), and (3) the common bases thesis (CB) are true.
(1) SCI: Categorical properties are not knowable by natural science.35
(2) CQ: Our qualitative experiences (i.e. qualia) are (solely or partially) constituted by their categorical bases (in some manner that is not [solely] via the causal, dispositional, and structural properties that are, in turn, borne by these bases).
(3) CB: Properties studied by science and our qualitative experiences share the same kind of categorical basis.
(See, for example, Chalmers 1996, 2003a; Stoljar 2001a, 2001b, 2006; Heil 2004a, 2004b; Seager 2006; Strawson 2008b; Alter & Nagasawa 2015)

35 Most Russellian Monists take (current) natural science to be a discipline that begins with causal measurements of external facts (Alter & Nagasawa 2015). Hence, facts we learn from methods that have no such beginning, such as introspection of first-person experience, presumably cannot be taken as parts of (current) natural science, regardless of whether they are justified knowledge. This is contrary to, say, some positivists’ view that natural science is all kinds of descriptions of our sensory content. I will follow this assumption in the rest of this section.
As Bertrand Russell (1992/1927a), Simon Blackburn (1990), Rae Langton (2015), and David Lewis (2009) suggest, in the case that CQ is true, categorical properties may – contrary to the claim of CI – be knowable through acquaintance with our qualitative experiences. Let us call this the *categorical acquaintance thesis*:

**The categorical acquaintance thesis:** (1) Through acquaintance with qualitative experiences, we can also acquaint ourselves with their categorical bases, and/or (2) through knowledge of qualitative experiences that is obtained through acquaintance, we can infer knowledge of their categorical bases.

The idea is very roughly that since some of our categorical bases are the constituents of our qualitative experiences, we can know them through our qualitative experiences. Of course, some metaphysical or epistemological assumptions have to be made in order to make sense of the thesis. I will discuss them as this dissertation goes on.

The Russellian Monist and the CI theorist both agree that SCI is true (i.e. categorical properties are not knowable by natural science). However, if the categorical acquaintance thesis is true, then CI is false, because we now have at least one method to know categorical properties, not – as CI tells us – none. As I have argued in Section 7, my reason for believing in CI is that, as far as I am concerned, there is no tenable combination of tenable hypotheses in natural science, metaphysics, and epistemology that allows us to know of categorical properties. However, the categorical acquaintance thesis, if tenable, will entail that my view is false. Hence, to show that my view is correct, I must show that the categorical acquaintance thesis is untenable.

In addition to the concern above, Russellian Monism has one more feature that makes it important. In the history of philosophy, many theorists holding other doctrines (such as idealism, Cartesian dualism, phenomenalism, Arthur Schopenhauer’s theory of the will, the Hindu theory of Brahman, and so on) have suggested their doctrines allow for acquaintance with categorical properties, or other entities or properties they take to be fundamental, but what is unique about Russellian Monism is that many influential theorists argue that it is a naturalistically respectable, or even physicalistically respectable, doctrine (Chalmers 1996, 2015; Stoljar 2001a, 2001b; Strawson 2013). (Recall that committing to physicalism or materialism entails committing to MN, while committing to MN does not entail committing to either physicalism or materialism. See the Introduction.) To understand the importance of this feature, we have to recall a default position of this dissertation,
which is that the prior probability of MN is assumed to be high (see the Introduction). Taking this into account, the categorical acquaintance thesis, in accordance with Russelian Monism, offers us two possibilities worthy of investigation.

The first is that, in the case that Russelian Monism is, indeed, naturalistically respectable, the categorical acquaintance thesis may rest on naturalistically respectable assumptions. This is important because if this is true, the objection to CI cannot be simply rejected because it rests on metaphysical non-naturalism, which our default position considers unlikely to be true (in the same way that many naturalist philosophers do not take seriously philosophical theses that rest on theism, idealism, Cartesian dualism, phenomenalism, Schopenhauer’s theory of the will, the Hindu theory of Brahman, etc.) In this case, the categorical acquaintance thesis might be used by the metaphysical naturalist as a response to CI.

The second is that, in the case that the argument that Russelian Monism is naturalistically respectable is not ultimately successful, a description of the world made by the Russelian Monist might still be shown by the argument to be very similar to one made by the metaphysical naturalist. With this in mind, most if not all conventional reasons in philosophy that contribute to the high likelihood of MN (say, the argument from physical causal closure in philosophy of mind, the problem of evil in philosophy of religion, the no-miracle argument in philosophy of science, and so forth) may apply equally to the likelihood of the Russelian Monist’s description of the world and cannot be used as counterevidence against it. With this in mind, if a serious inquiry into the nature of categorical properties turns out to favour Russelian Monism, it might be a significant challenge to MN.

In this section, I will first briefly introduce Bertrand Russell’s early formulation of Russelian Monism and the categorical acquaintance thesis. I will then introduce its recently developed versions that are claimed to be naturalistically or physicalistically respectable – in particular, the arguably most well-developed version, Daniel Stoljar’s Russelian non-standard physicalism. It is noteworthy that later Stoljar does not commit to the doctrine. Instead, he treats the doctrine as one that makes use of his favoured strategy for dissolving the mind/body problem in philosophy of mind. While he believes the strategy to be correct, he is increasingly less certain about which precise doctrine that makes use of the strategy is true. More details will be given as the discussion goes on. In addition, he does not believe that the categorical acquaintance thesis can rest on Russelian Monism. Nonetheless, I believe that Stoljar’s Russelian non-standard physicalism should be our target of investigation because it is the arguably best-developed version of Russelian Monism. Whether Stoljar personally commits to
the doctrine or not and whether he believes that the categorical acquaintance thesis can rest on Russellian Monism or not can be considered as independent issues.

8.2 Russell’s Russellian Monism

Let us begin with Bertrand Russell, an early proponent of the view in modern analytic philosophy, whom Russellian Monism is named for and inspired by. Despite the fact that Russell’s works are not contemporary, they are of genuine philosophical interest because of his interesting arguments for Russellian Monism and his defence of the categorical acquaintance thesis (which is rare among contemporary Russellian Monists). Contrary to many contemporary Russellian Monists who attempt to develop the thesis into a version of physicalism or panpsychism, Russell denies that fundamental ‘stuff’ (as he calls it) is mental or physical (1922/1921, p. 25); and he even explicitly denies being a materialist (1992/1927a, p. 382). What he is proposing is, instead, a version of neutral monism, which he takes to be inspired by William James (Russell 1922/1921, p. 25). Neutral monism, defined in Russell’s own terms, is as follows:

[B]oth mind and matter are composed of a neutral-stuff which, in isolation, is neither mental nor material. (1922/1921, p. 25)

By the terminology of contemporary philosophy of mind, I believe that the doctrine can be restated as follows:

Neutral monism: Mental and physical properties (or entities) metaphysically supervene on the same kind of properties (or entities), which are non-mental and non-physical.

Historical studies of Russell’s philosophy of mind show that Russell has kept on revising his philosophy of mind and has therefore held different versions of the doctrine throughout his whole life (Tully 2003, p. 332; Wishon 2015, p. 92). Nevertheless, since this section is not one on Russell’s neutral monism, but one on the doctrine bearing the title ‘Russellian Monism’ in contemporary literature, I shall focus on the version that contributes to Russellian Monism and set aside his other views. With this in mind, we shall discuss a view he expresses in his seminal book The Analysis of Matter (1992/1927a), which is widely considered to be the root of Russellian Monism (Stoljar 2001a, p. 269, 2006, p.107; Chalmers 2003a, p. 130; Alter & Nagasawa 2015, p. 424; Bennett manuscript).
8.2.1. The analysis of matter

In *The analysis of matter*, Russell proposes what he calls ‘psycho-cerebral parallelism’ (1992/1927a, p. 391). In his view, (1) physical properties are ‘causally dominant’, and (2) mental experiences are a part of the physical world and are ‘determined by the physical character of their stimuli’ (p. 391). To put this in the language of contemporary philosophy of mind and metaphysics, Russell believes that physical laws of causation are the only fundamental laws of causation; anything that is causal efficacious, including the mental, must causally act in ways that are (or supervene on) physical causation.

Nevertheless, for Russell, physics has its limits. Its aim is only ‘to discover what we may call the causal skeleton of the world’ (p. 391, my emphasis). It cannot tell us the ‘intrinsic character’ of matter. He describes the physical as follows:

> A piece of matter is a logical structure composed of events; the causal laws of the events concerned, and the abstract logical properties of their spatio-temporal relations, are more or less known, but their intrinsic character is not known. (p. 384)

Even though we cannot know of the intrinsic character of matter by physics, Russell believes that it can be known by our mental experiences. He writes:

> As regards the world in general, both physical and mental, everything that we know of its intrinsic character is derived from the mental side, and almost everything that we know of its causal laws is derived from the physical side. (p. 403)

In Russell’s view, it can be known by our mental experiences because (some of or a part of) such experiences *are* their intrinsic character. In a work published in the same year as *The Analysis of matter*, he writes:

> We now realise that we know nothing of the intrinsic quality of physical phenomena except when they happen to be sensations. (1927b, p. 154, my emphasis)

In sum, through physics we can only know about causal properties but not the intrinsic character of matter. But it can be safely assumed that mental experiences are not fundamentally different from other physical events, because the former and the latter share the same kind of causal nature, which
is their subjection to physical causation. Hence, they presumably share the same kind of intrinsic character. We can know about the intrinsic character of our mental experiences when we have them because (some of or a part of) our mental experiences are identical to their intrinsic character. Consequentially, we can also know about the intrinsic character of something physical, because mental experiences are, with respect to their causal nature, physical.

8.2.2. Is this Russellian Monism?

Recall that Russellian Monism is the following doctrine:

**Russellian Monism:** (1) The scientific categorical ignorance thesis (SCI), (2) the categorical qualia thesis (CQ), and (3) the common bases thesis (CB) are true.

(1) SCI: Categorical properties are not knowable by natural science.
(2) CQ: Our qualitative experiences (i.e. qualia) are (solely or partially) constituted by their categorical bases (in some manner that is not [solely] via the causal, dispositional, and structural properties that are, in turn, borne by these bases).
(3) CB: Properties studied by science and our qualitative experiences share the same kind of categorical basis.

I have introduced Russell’s theory in Section 8.2.2. As I have noted earlier, Russell’s theory and arguments are of genuine philosophical interest because of his interesting arguments for his theory (which I will discuss) and his defence of the acquaintance thesis (which is rare among contemporary Russellian Monists). However, an important matter of concern is whether his theory can be really considered as a version of the doctrine that contemporary theorists call Russellian Monism. This is important because what his thesis concerns is what he calls ‘the intrinsic character of matter’, and he does not use the term ‘categorical properties’. Hence, we have to determine whether what he calls the intrinsic character of matter and what contemporary Russellian Monists call ‘categorical properties’ are equivalent or analogous.

My approach to the issue is as follows. Since this dissertation is a work of contemporary analytic metaphysics, not history of philosophy, one way to read a historical philosophical text is to not be absolutely historically accurate. Other than reading the historical author sympathetically, we could also aim to search for a way of reading or an interpretation that is most tenable in terms of contemporary analytic metaphysics.
In line with the approach, I believe that Russell’s intrinsic character can be best interpreted as the kind of categorical property we are interested in. Let us begin with SCI. It is fairly obvious that Russell is committed to the doctrine. This is because he believes that physics offers us causal knowledge only, and that the intrinsic character is not causal. Since categorical properties are by definition non-causal properties, the intrinsic character must be a kind of categorical property. In addition, the kind of categorical property here is presumably not structural categorical properties such as molecular structures but fundamental categorical properties, because we obviously can know of properties like molecular structures (see a more detailed discussion in 4.3.3). Hence, it can be believed that Russell’s epistemology of properties is equivalent, or at least analogous, to SCI. In line with what has been said, if we substitute the term ‘categorical properties’ in CQ and CB, we can also discover that the doctrines are Russell’s views, which I discussed in Section 8.2.2.

However, there is a need to be careful when we interpret Russell’s view on acquaintance with intrinsic character in terms of Russellian Monism. While Russell says that we can know of the intrinsic character of our mental experiences when we are experiencing them, it is possible that the intrinsic character we know of is not any individual categorical basis of the mental experience but the aggregate of them. With this in mind, it is possible that we still do not know what a categorical property is like, because it is not obvious – or, at least, not uncontroversial – that we can always infer knowledge of a member from knowledge of an aggregate it belongs to. Nonetheless, what Russell provides is an example of what the categorical acquaintance thesis (or a very similar thesis) may be like. The idea is that we can know of some categorical properties of physical objects, regardless of whether they are aggregates or individuals, in our qualitative experiences. This is because they are the qualitative experiences, and a qualitative experience has no ontological status that is different from those of other physical objects but is, instead, a part of the physical world. Even though we are interested in categorical properties that are fundamental, the thesis may be something very close to what we want. More will be discussed in Section 10.

8.2.3. Russell’s argument for Russellian Monism

With the above overview of Russell’s theory, we wish to know what motivates him to advocate the theory. This is because one of the purposes of Part III is to determine whether the categorical acquaintance thesis (i.e. the view that we can know about our categorical bases by acquaintance) is successful; and for it to be successful, its assumptions – including Russellian Monism – have to be tenable. In addition, another of our purposes is to determine whether Russellian Monism, if is not
naturalistically respectable, is a formidable rival to MN or not. Due to the limit of space and the fact that this dissertation is not on Russellian studies, I will consider only one of his arguments, which I consider to be the most important. It is, however, noteworthy that Russell (as I read him) has other arguments that I, unfortunately, must set aside.

The argument is what I call the *argument from best fit*. The source of the argument is his book *An outline of philosophy* (1927b), which was published in the same year as *The analysis of matter*, so he is presumably holding the same view(s) in philosophy of mind. The text is as follows:

> Having realised the abstractness of what physics has to say, we no longer have any difficulty in fitting the visual sensation into the causal series. It used to be thought “mysterious” that purely physical phenomena should end in something mental. That was because people thought they knew a lot about physical phenomena, and were sure they differed in quality from mental phenomena. We now realise that we know nothing of the intrinsic quality of physical phenomena except when they happen to be sensations, and that therefore there is no reason to be surprised that some are sensations, or to suppose that the others are totally unlike sensations. The gap between mind and matter has been filled in, partly by new views on mind, but much more by the realisation that physics tells us nothing as to the intrinsic character of matter. (p. 154)

The idea, as I understand it, is an appeal to inference to the best explanation. My interpretation is as follows. We have difficulties in explaining mental phenomena. While we know that they happen in our brains, we, from the knowledge of physics, do not know where they can fit into our picture of the physical world. Meanwhile, on the one hand, Russell always emphasises that knowledge in physics consists of nothing intrinsic and is ‘abstract’, in the sense that it consists of only mathematical descriptions and predictions of changes of states (1927b, p. 150). On the other hand, fortunately for him, he already has the assumption from his own acquaintance that (some of or a part of) mental phenomena are intrinsic and non-abstract. As he puts it in *The analysis of matter* and *An outline of philosophy*:

> As regards the world in general, both physical and mental, everything that we know of its intrinsic character is derived from the mental side. (1992/1927a, p. 402)
Percepts are the only part of the physical world that we know otherwise than abstractly. (1992/1927a, p. 402)
We now realise that we know nothing of the intrinsic quality of physical phenomena except when they happen to be sensations. (1927b, p. 154)

With this in mind, we naturally fit mental phenomena, which are intrinsic, into the unknown intrinsic realm of the physical, like fitting a puzzle piece into what seems to be best hole. From this, we can finally have a complete picture of mental and physical.

Before assessing whether Russell’s argument from best fit is convincing or not (which I will do in Sections 10 and 12), I think it would be better to firstly assess recent developments of Russellian Monism, so that we can have an overall assessment of the doctrine.

8.3. Stoljar’s Russellian non-standard physicalism
Recall that one of our aims in Part III is to determine whether Russellian Monism is (physically respectable and hence) naturalistically respectable. Russellian Monism has many developed versions, and among them many claimed to be physically respectable and naturalistically respectable. Stoljar’s Russellian non-standard physicalism is arguably the most attractive one and is thus the one I shall focus on. Stoljar has never used the term ‘Russellian non-standard physicalism’. He has only used the terms ‘Russellian Monism’ and ‘non-standard physicalism’. But I combine the two here to avoid confusion because Stoljar has proposed another kind of non-standard physicalism, Nagelian Monism, according to which there is a yet unknown type of physical property that can explain our qualitative experiences. Stoljar considers Nagelian Monism to be a less specific doctrine that includes Russellian Monism as one of its many versions. While he always believes that Nagelian Monism is true, in his later works he is not certain which precise version of it is true.

As I have mentioned, Stoljar’s Russellian non-standard physicalism is not the only version of Russellian Monism that claims to be physically respectable and/or naturalistically respectable. David Chalmers and Galen Strawson have defended other versions (Chalmers 1996; Strawson 2008b). But their versions are both panpsychism, since Russellian intrinsic properties are believed to be protoconsciousness or experience, despite their claims that they are metaphysical naturalists.

---

36 It is noteworthy that Chalmers does not commit to Russellian Monism. Instead, he takes it as one of the few possibilities he finds attractive. Nevertheless, he considers the doctrine as the one he is ‘most drawn to’ (1996, p. 153).
(Chalmers 1996; Strawson 2013) and Strawson’s claim that he is a physicalist (2008b). On the other hand, Stoljar attempts to develop Russellian Monism into a version of non-panpsychistic physicalism (2001a, p. 273). This provides a much stronger connection to the version of MN I have developed in Part I, which is incompatible with panpsychism. (Recall that physicalism is a version of MN, and thus being a version of physicalism implies being a version of MN.) Hence, he will be the target of inquiry of this section after Russell.

The coming discussion will focus on how Stoljar develops Russell’s original thesis. I will also discuss the reason why Stoljar is increasingly less confident of it and eventually proposes his Nagelian non-standard physicalism as an alternative, because it will be useful to our assessment of Russellian Monism in the coming sections.

8.3.1. Agreements and disagreements between Stoljar and Russell’s Russellian Monism

I will argue that Stoljar’s Russellian non-standard physicalism and Russell’s Russellian neutral monism share very similar content. They do have many differences in their strategies to argue for their doctrines, their views on science and qualitatively experience that are relevant to the doctrines, and so forth. (In particular, Stoljar, in a personal correspondence, informed me that he rejects Russell’s categorical acquaintance thesis [p.c.].) But in terms of the matters that the two doctrines cover (not matters that are relevant to but not covered by the two doctrines), I will argue that they have no significant theoretical disagreements. Their main difference, as I will argue, rests on categorisation of ideas.

Like many contemporary writers, Stoljar focuses on categorical properties, not intrinsic properties or what Russell calls the ‘intrinsic character’ (2001a, 2006). But, as I have argued in Section 8.2.2, the two are closely related. In his very recent works on the (kind of) doctrine he calls non-standard physicalism, Stoljar replaces categorical properties with what he calls non-standard physical properties. The word ‘standard’ seems to stand for kinds of properties posited by our current physical theories, while the word ‘non-standard’ seems to stand for kinds of physical properties inaccessible to current physics. Nonetheless, while there can be many kinds of non-standard physical properties (because there can be many possible reasons a kind of property can be inaccessible to current physics), one of the possible distinctions between standard physical properties and non-standard physical properties he suggests still reflects the Russellian view. The distinction rests on ‘structural and non-structural, dispositional and non-dispositional, and so on’ (2015a, pp. 326-327). To avoid confusion with other non-standard physical properties, I will call the properties concerned Russellian non-
standard physical properties. After all, Russellian non-standard physical properties do not seem very
different from the kind of categorical property we are interested in. Recall that the kinds of properties
we are interested in are fundamental, intrinsic, and categorical, and they are by definition non-
dispositional, non-causal, non-structural, and so forth (see Section 4).

Stoljar, like Russell, believes that major problems in philosophy of mind surrounding qualitative
experience are results of the limit of physics (in its current form), and that taking categorical
properties as (a part of) the explanation of qualitative experience might provide a solution to the
problems. The exact details will be discussed in Section 8.3.3.

The most significant difference between Stoljar’s Russellian non-standard physicalism and
Russell’s neutral monism rests on whether the categorical properties should be considered to be
physical or neutral and hence whether Russellian Monism can be compatible with physicalism or not.
While Russell categorises categorical bases as neutral and explicitly denies their being physical,
Stoljar categorises them as non-standardly physical. But I think that this is a disagreement in
categorisation, not a theoretical disagreement. Our concern then should be why Stoljar believes that
he has provided a better categorisation than Russell has. This is because we wish to know whether
Russellian Monism is naturalistically respectable. (Recall that physicalistic respectability entails
naturalistic respectability.) Stoljar offers his argument for his categorisation in his influential article
‘Two Conceptions of the Physical’ (2001a), in which he argues for a conception of what he calls an
object-based conception of physicalism other than the common theory-based conception. Section
8.3.2 will be a detailed introduction to the two conceptions.

8.3.2. Two conceptions of the physical

Stoljar had not used the term ‘non-standard physicalism’ until recently, but the idea has always been
discussed since his early works. In his influential early article ‘Two conceptions of the physical’
(2001a) (and also in 2001b, 2015b), Stoljar argues that there are two kinds of physicalism: ‘t-
physicalism’ and ‘o-physicalism’. T-physicalism rests on the theory-based conception of the physical,
which is a conception held by most philosophers, including opponents of physicalism. Stoljar defines
the theory-based conception of the physical as the following:

[A] physical property is a property which either is the sort of property that physical theory
tells us about or else is a property which metaphysically (or logically) supervenes on the sort
of property that physical theory tells us about. (2001a, p. 256)
Characterising the physical as ‘the sort of property that physical theory tells us about’ seems to be the disciplinary characterisation of physicalism in the most direct and simple form I assessed in Section 2. Recall the definition I provided:

**The disciplinary characterisation (physicalism):** Every entity or property instantiated in the actual world is physical – in the sense of being posited by physics or being exclusively constituted by (i.e. reducible to or realised by) those posits.

Instead of taking reduction and reduction into concern like the above definition does, Stoljar’s definition takes supervenience into concern. But in the standard view in philosophy of mind, reductive and realisation relations are kinds of supervenience relations; supervenience is merely a more general concept than the concepts of reduction and realisation. In many cases, they are not considered very different.

Stoljar then suggests that there can be another kind of physicalism: o-physicalism, which makes use of an object-based conception of the physical. The conception is as follows:

A physical property is a property which either is the sort of property required by a complete account of the intrinsic nature of paradigmatic physical objects and their constituents or else is a property which metaphysically (or logically) supervenes on the sort of property required by a complete account of the intrinsic nature of paradigmatic physical objects and their constituents. (2001a, p. 257)

In this definition, the clause ‘the sort of property that physical theory tells us about’ in the theory-based conception is replaced by the clause ‘the sort of property required by a complete account of the intrinsic nature of paradigmatic physical objects and their constituents’. This concept of the physical is significantly similar to Jackson’s (1998) ostensive definition of the physical, which I assessed in Section 3.3.2.18. The methodology is to point out examples of physical objects, and then classify any property that serves the explanatory roles of it as physical. But Stoljar does not tell us what paradigmatic physical objects are. It is only hinted at by his following passage elsewhere:

---

37 Of course, this definition is one of physicalism, and the cited Stoljar’s text is, instead, a definition of the physical, but this is not a matter of concern. This is because this definition of physicalism still defines what the physical is; and Stoljar’s physicalism is still, much like this definition of physicalism, the view that there are only such properties in the actual world.
According to the object-based conception, for example if rocks, trees, planets and so on are paradigmatic physical objects, then the property of being a rock, tree or planet is a physical property. Similarly, if the property of having mass is required in a complete account of the intrinsic nature of physical objects and their constituents, then having mass is a physical property. (2015b)

The above passage seems to show that he means objects we intuitively consider to be physical or objects most of us agree to be physical (or something very similar).

Now the reason why categorical properties (which Russell considers neutral) can be, in Stoljar’s view, considered as physical should become clear. In Stoljar’s view, CI theorists such as Blackburn may be right in saying that physical theories cannot access categorical properties (2001a, p. 258), but as many philosophers – physicalists included – tell us, categorical properties are required to act as the bases of the properties we find in physics, such as dispositional properties (p. 259). With this in mind, categorical properties match the condition of being physical properties according to the object-based conception, which is to be ‘required in a complete account of the intrinsic nature of physical objects and their constituents’. In response to some claims of philosophers such as Chalmers (1996, p. 155) that Russellian Monism is a ‘non-materialist monism’,38 he argues that these philosophers ‘operate only with the theory-based conception’ and ‘miss the distinction between the theory-based and the object-based conception of the physical’ (p. 271).

The concern is then why we should prefer the object-based conception of the physical (and o-physicalism) over the theory-based conception of the physical (and t-physicalism). What Stoljar appeals to is an argument from cost (as I call it). In his work, he appeals to the unfavourable consequence of the theory-based conception. Consider unknowable categorical properties that many respected physicalists, such as Lewis, Armstrong, and Jackson, believe to exist. In his view, if the theory-based conception is correct, these properties cannot be considered to be physical because they are necessarily not within our physical theories. Physicalism would then be false because of the existence of the properties, which have nothing to do with philosophy of mind (p. 272). In a personal correspondence, he also added that Lewis, Armstrong, and many others would fail to be physicalists (p.c.). He then argues that, in order to avoid these unfavourable consequences, ‘it is necessary that [the physicalist] operates with a conception of the physical broader than simply the theory-based

38 Chalmers later comes to an agreement with Stoljar that Russellian Monism can be a version of physicalism (2015, p. 255).
conception’ (p. 272). In my understanding, his view is that the theory-based conception of the physical and o-physicalism are too narrow and thus costly, in the sense that they lead to some very unfavourable consequences, while the theory-based conception of the physical and t-physicalism are not and thus more preferable.

Let us return to the disagreement between Stoljar’s Russellian non-standard physicalism and Russell’s Russellian neutral monism. The disagreement rests on whether categorical properties (or something very similar) should be considered neutral or physical. In Stoljar’s view, most neutral monists, presumably including Russell, presuppose the theory-based conception of the physical because they define physical properties solely by physical theories and thus consider the supervenience bases of those physical properties as non-physical. However, once we distinguish the two conceptions of the physical, there is, in Stoljar’s words, ‘little reason’ to categorise the theory of concern as a version of neutral monism (p. 271).

An additional note that is important to the purpose of this dissertation. In response to the worry that his Russellian non-standard physicalism is a version of panpsychism, Stoljar argues that there is no reason to suppose so (p. 271). In his view, it merely follows the doctrine that some categorical properties will ‘in combination be the supervenience base for qualitative properties’ (pp. 272-273), but they themselves may be non-qualitative.

8.3.3. Stoljar’s arguments for Russellian non-standard physicalism

Recall that Russellian Monism (as I understand it) consists of three doctrines:

(1) **The scientific categorical ignorance thesis (SCI):** Categorical properties are not knowable by natural science.

(2) **The categorical qualia thesis (CQ):** Our qualitative experiences (i.e. qualia) are (solely or partially) constituted by their categorical bases (in some manner that is not solely via the causal, dispositional, and structural properties that are, in turn, borne by these bases).

(3) **The common bases thesis (CB):** Properties studied by science and qualitative experiences share the same kind of categorical basis.

Recall two purposes of Part III. One is to determine whether the categorical acquaintance thesis is successful. Even though Stoljar, as he informed me in a personal correspondence (p.c.), does not accept the thesis, his arguments for Russellian Monism are worthy of consideration because
Russellian Monism is what the thesis rests on. Another purpose is to determine whether Russellian Monism, if is not naturalistically respectable, may be a formidable rival to MN. Even though Stoljar considers it to be physicalistically respectable (which implies its being naturalistically respectable), his arguments for the formidability of Russellian Monism are worthy of independent consideration.

Stoljar never explicitly commits to SCI, even though he seems to consider Blackburn’s argument for it to be attractive (2001a, p. 258) and has co-written an article with Michael Smith (1998) that argues that SCI is a necessary consequence of an epistemological thesis. I shall not go further into this issue because I believe we already had sufficient discussion of issues concerning CI in Section 7.

On the other hand, CB is an assumption Stoljar does not – and need not – argue for. This is because it is a consequence of physicalism, which he assumes to be true. Note that if the doctrine is false (i.e. qualia have a kind of categorical basis that is different from those of the properties studied by science), this will imply that dualism is true, which is inconsistent with physicalism. Hence, I shall set SCI and CB aside and focus on his arguments for CQ.

In his ‘Two conceptions of the physical’ (2001a), which we have been focusing on, Stoljar seems to be a proponent of CQ. But in another article of the same year, ‘The conceivability argument and two conceptions of the physical’ (2001b), he already keeps a distance from the doctrine, even though he obviously considers it to be attractive (2001b, p. 48). And in his later writings, he eventually becomes less confident of it. In this sub-section, I will discuss Stoljar’s arguments for CQ; and Stoljar’s reason to be less confident will be left to Section 8.3.4.

Stoljar takes CQ to be attractive because he believes the doctrine can provide an ‘elegant’ solution to two major problems in philosophy of mind concerning qualia (2014, p. 19): the knowledge argument (2001a) and the conceivability argument (2001b). As he writes in a later article:

The Russellian monist, as I understand matters, is someone who agrees that materialism (at least of the usual sort) is defeated by the conceivability argument but holds out the hope for some alternative account that preserves its spirit and structure and so avoids the inelegance of dualism. (2014, p. 19)

*The knowledge argument* is roughly the following. If physicalism is correct that everything is physical, omniscient physical knowledge would entail omniscient knowledge of everything, including knowledge of all mental states. But omniscient physical knowledge does not entail what qualia – for example, the experience of seeing redness – is like. The most famous version of this
argument is Jackson’s story (1982, 1986) about a perfect physical scientist, Mary. She has spent her whole life living in a black and white room. Jackson argues that she will learn something new when she is released and sees redness for the first time. The conclusion is that qualia cannot be physical, because knowledge of qualia is not physical knowledge.

Stoljar’s solution is that omniscient physical knowledge does not entail knowledge of qualia when it is t-physical. Mary’s perfect physical knowledge is t-physical because she is merely perfect in her knowledge of physical theory. But Mary’s physical knowledge is not perfect in terms of o-physical properties. She cannot know about categorical properties no matter how much she knows about physical theory (2001a, p. 264). Nevertheless, physicalism can be true because all mental truths are entailed by o-physical truths.

*The conceivability argument* is roughly the following. Chalmers (1996) famously argues that philosophical zombies, i.e. perfect physical duplicates of us lacking qualia, are conceivable. In Chalmers’s view, if something is conceivable and so it is metaphysically possible, and hence zombies are metaphysically possible. But physicalism is inconsistent with the metaphysical possibility. If physicalism is true, perfect physical duplicates of us will necessarily be perfect mental duplicates of us, and zombies would thus be metaphysically impossible. Therefore, physicalism is false. Qualia must be entities or properties somehow metaphysically independent of the physical and thus are not physical.

Stoljar’s solution is to argue that o-physical zombies are ‘weakly-conceivable’ (Stoljar 2001b, p. 400). This is because we do not really know of our categorical bases. Therefore, we cannot. Stoljar seems to be suggesting that, if we actually have no idea of the nature of the conceived things, our conceivability is not strong enough to meet the requirement of the argument, in the sense that we can genuinely tell whether such zombies have qualia. In this case, ignorance has defeated the inference from conceivability to metaphysical possibility. Zombies are only untruthfully conceived to be possible because we do not have sufficient knowledge.39

It can be said that Stoljar’s argument for Russellian non-standard physicalism is very similar to Russell’s argument from best fit. Russellian non-standard physicalism is preferred because it can fit qualia into the picture of physicalism and solve several difficulties in a good (or, in his word, elegant)

39 This is, no doubt, inconsistent with Russell’s view. In Russell’s view, our mental experiences are explained by our intrinsic character, about which we can know in the experiences. With this in mind, if something shares our intrinsic character and physical causal structure, they must also share all of our mental properties. Hence, o-physical zombies are presumably inconceivable. However, this is not problematic since Stoljar, as he informed me in a private correspondence, does not share Russell’s view on categorical acquaintance (p.c.).
way. But he does not consider the doctrine to be the best fit, merely a good fit. This is probably because he does not assume that we can know of categorical properties or anything similar through acquaintance with qualitative experiences like Russell does, and hence, in his case, the connection between qualia and categorical properties can be said to be a lot more speculative. As he puts it in his ‘Two conceptions of the physical’:

Of course, even if our argument is right, it has only been shown that this is one way of resolving the debate. It has not been shown that this is the best or the only way. (2001a, p. 277)

This is why Stoljar has been continuing to explore other possible good fits and is increasingly less confident of Russellian non-standard physicalism. In another article of the same year as ‘Two conceptions of the physical’, ‘The conceivability argument and two conceptions of the physical’ (2001b), Stoljar already claims that ‘our account does not require the Kantian picture’ (2001b, p. 48). Recently, he has developed what he calls ‘Nagelian Monism’ (Stoljar 2015a), which I will elaborate on in the next section.

8.3.3. Stoljar’s reason for doubting Russellian non-standard physicalism

I am ending this my introduction to Stoljar’s views with his reason for doubting Russellian non-standard physicalism. This is because I consider it to be useful to the assessment of Russellian Monism in the coming sections in Part III. Recall that we need to determine whether the categorical acquaintance thesis is successful or not. If Russellian Monism is not true, then any theory that takes it to be an assumption will face challenges.

Stoljar eventually changes his mind mainly because he takes Russellian non-standard physicalism to be one of the several attractive views that takes a kind of yet unknown physical property to be (a part of) the explanation of qualia. While he is confident that there is such a type of yet unknown physical property, he has never been very certain what it is. As more possibilities of what such a kind of property might be have been explored in his works, Russellian non-standard physicalism no longer has such a crucial role in Stoljar’s thought as it used to.

In an article of the same year as ‘Two conceptions of the physical’, ‘The conceivability argument and two conceptions of the physical’ (2001b), Stoljar already claims that ‘[his] account does not require the Kantian picture [i.e. CI]’ (p. 408). Although he is still focusing on and defending
Russellian non-standard physicalism in that article, he argues that it is not the truth of the model of Russellian non-standard physicalism that matters. Russellian non-standard physicalism only serves the purpose of demonstrating what ‘ignorance of certain features of the physical world’ is like (p. 408). It is only an example of such ignorance, not the only way of having such ignorance. For Stoljar, what is important is that such ignorance may exist, and that such ignorance can respond to the conceivable argument and the knowledge argument. He calls it the missing concept hypothesis (p. 409).

A more developed version of the missing concept hypothesis is seen in Stoljar’s 2006 book Ignorance and Imagination. In the book, he proposes the ignorance hypothesis, according to which ‘we are ignorant of a type of experience-relevant nonexperiential truth’ (p. 6, my emphasis). Stoljar’s strategy is not to prove that the ignorance hypothesis is true, but to show that it is an ‘an extremely reasonable hypothesis to hold’ by demonstrating some possibilities of ignorance (p. 141). Three such possibilities are discussed in detail. (i) Russellian non-standard physicalism is one of them. The other two are (ii) an objection against excessive optimism of our epistemic power and (iii) historical precedent. The former is the hypothesis that we do not have an ideal epistemic grasp of the world (p. 88). To be more precise, it is (early) Jackson’s (1982) hypothesis that our cognitive powers might possibly be limited because of the way we evolved. The latter is the hypothesis that we might, much like many cases of our ancestors, face an explanatory gap because some type(s) of forthcoming scientific knowledge are currently still lacking. Stoljar argues that there are many such historical examples in the history of science. In particular, Stoljar discusses C. D. Broad’s emergentism of chemistry in detail. Broad argues that there is an explanatory gap between water, and oxygen and hydrogen atoms, and hence materialism is false. But it is now obvious that ‘Broad was ignorant of a type of nonchemical truth relevant to the nature of chemistry’ (p. 140). I end this very brief introduction to these two possibilities here because it is irrelevant to my purpose, which is to inquire into Russellian Monism.

In an article in 2015, ‘Russellian Monism or Nagelian Monism?’ (2015a), Stoljar develops the ignorance hypothesis into what he calls ‘Nagelian Monism’, though in an earlier article in 2014, ‘Four kinds of Russellian Monism’ (2014), he has already proposed the same view with name ‘RM4’, which stands for ‘Russellian Monism 4’ (despite the fact that the view is not Russellian). Contrary to Russellian Monism, Nagelian Monism ‘makes no theoretical use of the dispositional/non-dispositional distinction […], the role/realiser distinction […] and the structural/non-structural distinction’ (2014, p. 36). This view is called Nagelian Monism (or [Nagelian] non-standard
physicalism) because it is inspired by the following passage in Thomas Nagel’s *The view from nowhere* (1986):

The difference between the mental and the physical is far greater than the difference between the electrical and the mechanical. We need entirely new intellectual tools, and it is precisely by reflection on what appears impossible – like the generation of mind out of the recombination of matter – that we will be forced to create such tools. It may be that the eventual result of such exploration will be a new unity that is not reductionist. We and all other creatures with minds seem to be composed of the same materials as everything else in the universe. So any fundamental discoveries we make about how it is that we have mind and what they actually are, will reveal something fundamental about the constituents of the universe as a whole. (Citation of Nagel in Stoljar 2014, pp. 31-32)

For the Nagelian Monist, the Russellian standard/non-standard distinction of physical properties is now replaced by another set of standard/non-standard distinction. Other than the dispositional/non-dispositional distinction of Russellian Monism, it is a distinction between being described in contemporary physics and being described in ideal physics. Stoljar writes:

For the Nagelian, standard physical properties are properties of the sort described in contemporary physics (and perhaps contemporary empirical inquiry generally). Correlatively, non-standard physical properties are properties of the sort described in the physics, whatever it is that one will or might formulate at the ideal limit of inquiry. (2015a, pp. 327-328)

In terms of being versions of ‘the missing concept hypothesis’, Nagelian Monism is a much less specific doctrine than Russellian Monism. The view is that we are ignorant of (at least) a type of experience-relevant nonexperiential truth, but we do not know what it is.\(^{40}\)

It is noteworthy that Nagelian Monism, as a non-standard physicalism, is a t-physicalism that rests on a theory-based conception of the physical, because it is defined with reference to the ideal

\(^{40}\) It might be useful to some readers to note that Stoljar’s Nagelian Monism is inspired by Nagel’s text in *The view from nowhere*, which I cited above, but is not actually Nagel’s view in the book and is, in fact, in conflict with it. In the book, Nagel also notes that the type of experience-relevant truth concerned must be, in his words, ‘subjective’ (2014, p. 26). This is in opposition to Stoljar’s view that it is nonexperiential.
physical theory. It is unclear how it can be consistent with Russellian non-standard physicalism, which Stoljar takes to be a version of o-physicalism. (Recall that physicalism that rests on the object-based conception of the physical.) This is a bit puzzling, since Stoljar considers Russellian Monism to be a version of Nagelian Monism. No doubt, he may consider the ignorance of categorical properties simply as a limit of current physics, not one of ideal physics; Russell, indeed, has this view (1997/1935, p. 204). But even so, it hardly seems to be something we know for sure, and I doubt that Stoljar would accept the fact that categorical properties are necessarily non-physical simply because it turns out to be false. (Recall his argument from cost that I discussed earlier in Section 8.3.2, by which he argues that we should not accept a definition of physicalism that will imply the falsehood of physicalism because of metaphysical assumptions that have nothing to do with philosophy of mind, and will count Lewis, Armstrong, and many others as non-physicalists.) However, in any case, since the consistency of Stoljar’s definitions of physicalism is not a matter of concern of this section, I will set it aside.

Why does Stoljar prefer Nagelian Monism over Russellian Monism? One reason is that he believes what he calls the from-structure-only-structure thesis (i.e. the thesis that structural, relational, or dispositional truths can only deduce further structural, relational, or dispositional truth) can hardly be shown to be true (2006, p. 151). Hence, even if qualia are non-causal, non-dispositional and non-structural in nature, it is difficult to see why (non-structural) categorical properties are required to account for them. This is because without the from-structure-only-structure thesis, it is still possible that they are realised by causal, dispositional, or structural properties.

One last note before I end my introduction to Stoljar’s views. It should be noted that Stoljar, as a Nagelian non-standard physicalist, still seems to consider Russellian non-standard physicalism to be highly likely to be true. Russellian non-standard physicalism is among the three possibilities of ignorance that Stoljar considers at length. In comparison, there are other possible cases, like ignorance of neurophysiology, that he very briefly mentions but offers no consideration (2006, p. 71). Therefore, I believe it is fair to say that Stoljar’s position on the issue of Russellian non-standard physicalism is as follows: Nagelian non-standard physicalism is true; and if Nagelian non-standard physicalism is true, Russellian non-standard physicalism is highly likely to be true.

8.4. Conclusion

The main thoughts of Russell and Stoljar have been sketched above. We can return to the purpose of Part III in the coming sections on the basis of the theoretical frameworks Russell and Stoljar offer,
which is to answer three questions: (1) Is the categorical acquaintance thesis successful? (2) Is Russellian Monism an attractive doctrine? (3) Can Russellian Monism can be a version of or a doctrine similar to (physicalism and hence) MN? Recall that the importance of the questions is as follows: (I) If the categorical acquaintance hypothesis is true, CI is false. (II) If Russellian Monism can be a version of MN, then the categorical acquaintance thesis can be grounded on naturalistically respectable assumptions, which may allow it to resist the possible objection that it is not naturalistically respectable. (III) If Russellian Monism cannot be a version of MN, it may still be an attractive doctrine and be similar enough to MN, which allows it to avoid most shortcomings of other versions of metaphysical non-naturalism and thus be the most formidable rival of MN.
Part III

Section 9. The categorical acquaintance thesis

9.1. Introduction

In Section 8, I sketched Russell and Stoljar’s Russellian Monism (RM). In this section, I will return to the purposes of Part III, among which is to answer the following two questions. (1) Is the categorical acquaintance thesis (henceforth, CA) true? (2) Is RM likely to be true? Question (1) is important because if the CA is true, the categorical ignorance thesis (CI) is false; question (2) is important because RM is the metaphysical framework CA takes to be an assumption.

The structure of this section is as follows. In Section 9.2, I will introduce the relation between RM and CA. In Sections 9.3 and 9.4, I will assess how CA can actually be developed from the assumptions of RM in the most tenable way possible. In Section 9.5, I will argue that RM, with the categorical acquaintance hypothesis or not, has a paradoxical epistemology of qualia, which does not allow for any belief or memory about qualia, and hence both the categorical acquaintance hypothesis and RM are implausible.

9.2. RM and categorical acquaintance

Recall that RM and the categorical acquaintance thesis (henceforth, CA) are the following doctrines:

**RM:** (1) The scientific categorical ignorance thesis (SCI), (2) the categorical qualia thesis (CQ) and (3) the common bases thesis (CB) are true.

1. **SCI:** Categorical properties are not knowable by natural science.
2. **CQ:** Our qualitative experiences (i.e. qualia) are (solely or partially) constituted by their categorical bases (in some manner that is not [solely] via the causal, dispositional, and structural properties that are, in turn, borne by these bases).
3. **CB:** Properties studied by science and our qualitative experiences share the same kind of categorical basis.

The categorical acquaintance thesis (CA):
(CA1) Through acquaintance with qualitative experiences, we can also acquaint with their categorical bases, and/or

(CA2) through knowledge of qualitative experiences that is obtained through acquaintance, we can infer knowledge of their categorical bases.

Several influential CI theorists, including Simon Blackburn (1990), Rae Langton (2015) and David Lewis (2009), believe that CA is a formidable objection to CI. In Section 9, we have also seen that Russell says that we can know of the intrinsic character of our mental experiences when we are having the experiences, because (some of or a part of) such experiences are their intrinsic character. In Section 9.2.2, I have argued that the intrinsic character Russell has in mind is best understood as categorical properties. Unfortunately, Russell does not explain in what precise way CA can work.

It is important to note that not all agree that CA is consistent with RM. Some writers argue that acquaintance with qualia is one thing and acquaintance with their categorical bases is another, and hence what their categorical bases are like remains a mystery (see, for example, Robinson 2009, pp. 197-198; Majeed manuscript a).

With this in mind, I will critically assess several possible ways CA can work in Sections 9.3 and 9.4. In particular, I will argue that a particular version of CA2 has a chance of success. It is noteworthy that the assessment will be done with the assumption that RM is correct and arguments against it have no force, even though I will argue for the contrary in Section 9.5.

9.3. CA1

In an article co-written with Christopher Robichaud (2010), Langton discusses RM and writes:

If [psychological] properties are intrinsic (there is room for doubt here), then there are some fundamental intrinsic qualities with which we have acquaintance. (p. 175)

Despite the fact that they are sceptical about the truth of the proposal, they go as far as calling it a ‘surprising solution to the problem of humility’ and ‘an exciting proposal’ (p. 175). Unfortunately, they, like Russell, do not spell out in detail how the proposal can actually work. I, in disagreement with them, believe that it is in fact a pretty puzzling proposal. This is because it is very difficult to see why acquaintance with categorical properties would follow from acquaintance with qualia.
Some might think of simply applying Leibniz’s Law, according to which if \( x \) and \( y \) are identical, then \( x \) has every property of \( y \) and \textit{vice versa}, to our qualia and their categorical bases, so that the property of being acquainted with can be shared among the them because of their identity.\(^{41}\) But we must be careful when we apply the law because it must be made clear whether the property concerned is intentional or not. If it is intentional because it is a matter of, say, desire, perception, knowledge, or likewise, then there is a consensus within contemporary philosophy that Leibniz’s Law cannot apply, because of the limits of human psychology and knowledge. For example, as J. J. C. Smart (1959) famously teaches us, one may have knowledge of, say, lightning even though she knows nothing about electricity, despite the fact that lightning is identical to an electrical phenomenon. No doubt, in this case she may still be said to have some knowledge about electricity in some \textit{de facto} sense, but it is not something she is aware of.

But we must be careful when we apply the law because it must be made clear whether the property concerned is intentional or not. If it is intentional because it is a matter of, say, desire, perception, knowledge, or likewise, then there is a consensus within contemporary philosophy that Leibniz’s Law cannot apply, because of the limits of human psychology and knowledge. For example, as J. J. C. Smart (1959) famously teaches us, one may have knowledge of, say, lightning even though she knows nothing about electricity, despite the fact that lightning is identical to an electrical phenomenon. No doubt, in this case she may still be said to have some knowledge about electricity in some \textit{de facto} sense, but it is not something she is aware of.

Contrary to the worry above, perhaps acquaintance is not intentional like \textit{knowledge} but is, instead, unintentional. It might be, I speculate, certain capabilities to know of something. Hence, by the way Leibniz’s Law works, one may have certain capabilities to know of the categorical bases of her own qualia in the same way that she has certain capabilities to know of the qualia. Consider the case that my hard copy of David Armstrong’s \textit{A materialist theory of the mind} is identical to some pieces of paper. Now I may form a knowledge-generating relation with the book (e.g. weighing it

\(^{41}\) This might be potentially confusing is because the kind of categorical property we are interested in are fundamental properties. To be more precise, considering ‘being acquainted with’ as a property implies that it is considered as a property possessed by a property; and it is not obvious that fundamental properties can possess properties. To avoid such a confusion, it might be useful to remind that, as it is agreed among philosophers, there are two different meanings of the term ‘properties’ in the literature: it might mean constituents of objects; and it might also mean predicates. Throughout this dissertation, we have been concerning properties in its former meaning, while the kind of property Leibniz’s Law concerns is the latter.
with a scale). No doubt, it implies that I am in the same relation with the pieces of paper of which the book is made. This case also allows us to be correct about our own qualia but mistaken about our categorical bases. One may have used her capabilities to know of her own qualia properly without doing the same to their categorical bases. This is because she may have not used her capabilities to know of the categorical bases properly, or may not be aware of the fact that she has such capabilities with regard to them. After all, it is not difficult to imagine that someone correctly weighs *A materialist theory of the mind* but is still mistaken about the weight of the pieces of paper that the book is made of.

This proposal, I believe, cannot be said to be mistaken, but it cannot offer us what we really want. The proponent misses the point that even if qualia and their categorical bases are identical, strictly speaking they are identical to the *aggregate of their categorical bases*, not any of their *individual categorical bases*. Recall our example of weighing *A materialist theory of the mind* and the pieces of paper that the book is made of. Even if my capability to weigh *A materialist theory of the mind* implies my capability to weigh the pieces of paper that the book is made of, the question is whether I therefore have a capability to know about the weight of any single piece of paper among them. The answer is, obviously, ‘No!’ In sum, I believe that in order to make sense of CA1, the proponent needs a more sophisticated epistemological thesis than a simple identity claim with respect to qualia and their categorical bases.

**9.3.2. The identification thesis**

In his article on CI, ‘Ramseyan humility’ (2009), Lewis takes RM seriously (p. 217) and argues that the doctrine may provide a response to CI when it is in combination with what he calls the *identification thesis*. As he puts it:

[The Russellian Monist] might also accept the Identification Thesis: anyone acquainted with a quale knows just which property it is. Now it may seem that we can know the identities of the fundamental properties after all—we need only become acquainted with them. (p. 217)

Elsewhere, Lewis provides more detail of the thesis:

I say that according to the Identification Thesis, the knowledge I gain by having an experience with quale $Q$ enables me to know what $Q$ is – identifies $Q$ – in this sense: any
possibility not ruled out by the content of my knowledge is one in which it is $Q$, and not any
other property instead, that is the quale of my experience. Equivalently, when I have an
experience with quale $Q$, the knowledge I thereby gain reveals the essence of $Q$: a property
of $Q$ such that, necessarily, $Q$ has it and nothing else does. (1995, p. 142)

The idea is that simply by experiencing a quale, we will come to be in a unique – and, indeed,
idosyncratic – epistemic position. In the position, we can know about all the essences of the quale.
As Lewis notes, the extent we can know about the quale according to the thesis is ‘uncommonly
demanding’ (1995, p. 141). In his view, the thesis is strong enough to enable us to know about the
constitution of our qualia, and is thus inconsistent with physicalism because of the fact that we cannot
know the neurophysiological structure of a quale simply by experiencing it (1995, p. 142). By the
same reasoning, he believes that if RM and the thesis are both true, then we are in a position to know
of the categorical bases of a quale simply by having one.

Lewis himself does not believe in the thesis. His response to the thesis is simple: he rejects it
simply because he is a proponent of materialism, which he takes to be inconsistent with it (1995, p.
142; 2009, p. 217). For our purpose here, the matter of concern is why Lewis takes it so seriously as
a possibility. The reason Lewis offers us is that it is a part of our folk psychology (1995, p. 141). He
believes so because of his observation of other philosophers:

Why do I think it must be part of the folk theory of qualia? Because so many philosophers
find it so very obvious. I think it seems obvious because it is built into folk psychology.
Others will think it gets built into folk psychology because it is so obvious; but either way,
the obviousness and the folk-psychological status go together. (1995, p. 142)

There are a lot of contestable assumptions made in the cited passage, such as the assumption that
philosophers (whose folk intuitions may be much disrupted by their sophisticated philosophical views)
can be taken as good samplings in a study of folk psychology, the assumption that there are really
many philosophers who find the thesis obvious, and so on; and there are also some sceptical questions
that can be asked, such as whether a weaker version of the thesis can explain the same thing, whether
folk psychology (or the particular kind of folk psychology the thesis belongs to) is reliable enough to
be worthy of being taken seriously, and so forth. No doubt, because the matter of concern is what our
folk psychology is like and how we can collect evidence about it, those assumptions and sceptical
questions are all empirical matters and thus require independent inquiries. Nevertheless, my point is that many of Lewis’s assumptions are fairly contestable, and hence it is not very obvious that we should assume the identification thesis ought to be taken seriously like Lewis does.42

Regardless of whether the thesis is a part of folk psychology or not, I believe that its fatal shortcoming is that it is obviously false. Let us perform a simple experiment. The reader can look at this whole page without paying much attention to what I say. Can the reader really tell what each letter looks like, despite the fact that he or she has each letter as part of his or her visual image? For most of us, the answer will be ‘no’. Consider also Lewis’s famous example of the quale of tasting vegemite. Can every lover of vegemite tell exactly which kinds of tastes (such as the tastes of saltiness, bitterness, and so forth) play a role? The answer, again, seems to be ‘No’. In sum, it seems to be an obvious – and even demonstrable – fact that we often cannot know the nature and/or elements of our qualitative experiences. Hence, it seems that a realistic phenomenology that takes our actual experiences – rather than some very abstract intuitions we have with regard to metaphysics (and/or epistemology) – into concern can hardly be consistent with the identification thesis.43

In addition, the thesis seems to be in conflict with RM. If the thesis is, as Lewis says, inconsistent with standard physicalism, then it should be not consistent with RM either. This is because we certainly do not come to know that our qualia consist of their categorical bases just by experiencing and knowing them. One might suggest that primitivism, according to which at least some of or some elements of our qualitative experiences are fundamental properties, might provide a solution to the problem (see Majeed manuscript b). This is because this doctrine allows for the view that what we know about, say, the experience of redness already include all possible knowledge about it. But a combination of the doctrine and RM seems to be implausible. Provided that the number of (elements of) our qualitative experiences we can have is large, then – according to the view – we will have a tremendously large number of kinds of categorical bases. But we know that our brain is (probably) only made up of a small number of kinds of fundamental particles. In addition, neuroscience teaches that a lot of our experiences share common supervenience bases. It seems very difficult to make sense of the view that each (element of our) qualitative experience has an independent categorical basis.

42 For detailed discussions, see Stoljar 2006, 2009.
43 I am heavy indebted to a presentation by Andrew Lee (2016) with respect to these considerations.
9.4. CA2

Recall that CA2 is the thesis that ‘through knowledge of qualitative experiences that is obtained through acquaintance, we can infer knowledge of their categorical bases’. I am a bit more optimistic about this thesis because the diversity of knowledge inference may allow there to be more interesting possibilities of how it might work. Here I shall focus on inference to the best explanation (IBE), which, in combination with some views on meditation, I believe to have a chance of success if other problems with RM (which include some I believe to be fatal) are set aside.

9.4.1. A quick note on deductive inference

I shall focus on IBE because deductive inference obviously does not work. By a *deductive inference*, I mean the process to infer facts by the method of logical deduction. Recall my example of weighing David Armstrong’s *A materialist theory of the mind* in Section 9.3.1. No doubt, we can have a deductive inference from the fact that the aggregate of the pieces of paper is 1.1 pound heavy to the fact that a single piece of paper among it is not more than 1.1 pound. This is, indeed, a very substantive piece of information, but is also very imprecise. We have seen in Section 7 that what CI targets is precise knowledge, not imprecise knowledge.

9.4.2. *Inference to the best explanation (IBE)*

IBE, on the contrary, does not solely rely on deductive logic. It works when we take it to be the case that there is a best explanation – often by speculations – that can explain certain facts. The method is widely used in science, since a hypothesis rarely logically follows from observations. By IBE, we may speculate about what our categorical bases are like by considering which possible form of a categorical property can explain what our qualia are like.

As I have argued, our qualia cannot plausibly be any of their individual categorical bases. They may instead be aggregates of their categorical bases or some kind of higher-level property or entity that supervenes on their categorical bases. But by the method of IBE, we may speculate about what a categorical basis is like by the knowledge we have about qualia, even if we do not have a direct acquaintance of it.

Nevertheless, to make use of our qualia to speculate about what their categorical bases are like still seems to be a difficult job. Given the fact that we have a tremendously wide variety of qualitative experiences, it is difficult to infer from them what a categorical basis is like. This is because of a
dilemma as follows: we either use all of our qualitative experiences or a particular kind (or element) of qualitative experience as the target to be explained. In the former case, different kinds (or elements) of our qualitative experiences (such as phenomenal colours, phenomenal sounds, smells, tastes, and so on) seem to have little in common, and hence it is difficult to see how anything can be inferred from them. In the latter case, we may pick up an itch, for example, as a sample and infer that a categorical basis is like a very slight itch, but we lack a justifiable reason to choose it among others, among which each may lead to significantly different results.

9.4.3. IBE and meditation

I believe the solution to the dilemma caused by our wide variety of qualitative experiences can be solved by finding out a kind of qualitative experience that is found to be the simplest or to be basic. This is because its lack of complex content may allow it to have the closest resemblance to its categorical bases, which are presumably also simple, and thus allow us to infer what our categorical bases are like.

Some extraordinary meditative experiences seem to be able to provide such a possibility. According to many accounts and reports, they have a difference in nature from our ordinary experiences. Precisely, they are often said to be simple, basic (in the sense that they lack representational contents, emotional contents, thinking, and so on), and even ‘transcendental’ or ‘most real’ (which are difficult to explain) (see, for example, Fontana 2007; Fasching 2008). Even though it is not easy to determine how probable it is that we can infer from meditative experiences what a categorical basis is like, there seems to be at least a chance of success. Indeed, many religious traditions consider meditative experiences to be a means of knowing what they believe to be the fundamental entity or property. Some Hindu and Buddhist mystics even precisely proclaim that extraordinary meditative experiences allow us to see through causation and relations and grasp the ultimate reality, which is non-causal and non-relational. What they call the ultimate reality might be equivalent or somehow analogous to categories properties, because both are fundamental, non-causal, and non-relational.

This, of course, cannot solve all problems surrounding our wide variety of qualitative experiences. In particular, there is a problem that David Chalmers (forthcoming) calls the palette problem and that Karen Bennett (manuscript) discusses in length. It is roughly the following. Given our wide variety of qualitative experiences, how can they be constituted by ‘microqualities’ or ‘protoqualia’ (as the authors call them), which variety is presumably small (regardless of whether
these microqualities or protoqualia are the categorical bases of our qualia or, instead, some properties or entities at the intermediate level between our qualia and their categorical bases)? Yet this problem is not particular to choosing meditative experiences as a starting point of inference but a problem facing any tenable version of RM. This is because there is always the problem of how our limited amount of kinds of categorical bases can constitute our diverse qualitative experiences. I will leave the discussion of the problem to Section 10 and set it aside for the moment. At least, if the general problems facing RM are set aside, meditative experiences seem to be an attractive starting point of IBE.

Nevertheless, I believe that there are at least some questions that any philosopher who is attempting to solve CI by the method must answer. The first question is, of course, whether the particular kind (or interpretation) of account and report that can support its use here is reliable. In particular, as David Hume (1966/1777) famously reminds us, when a claim has extraordinary content, then the likelihood that there are ordinary mistakes involved might be even higher than the likelihood that it is true. I shall not discuss this in detail because this has to be left for empirical inquiries.

The second question concerns the simplicity and basicity of meditative experiences. In response to some recent hypotheses in cognitive science and neuroscience that consciousness can be measured in levels, Tim Bayne, Jakob Hohwy, and Adrian Owen (2016) express their worry that consciousness might be multidimensional. The idea is that each state of consciousness might have multiple levels with respect to different dimensions. I believe that the same problem arises here when meditative experiences are considered to be simple and basic. While meditative experiences are, according to many accounts and reports, simple and basic, it is unclear whether they are qualitatively simple and basic. They might be, instead, functionally simple or basic. Even if they are qualitatively simple and basic, levels of qualitative experiences might be multidimensional as well. Hence, they, again, may be simple or basic in one respect but not in another. In fact, this possibility may contribute to the fact that accounts and reports of meditative experiences are diverse, and that not all extraordinary meditative experiences are reported to be simple and/or basic (Fontana 2007). The explanation is that different practitioners may focus on different dimensions of meditative experiences.

The last question is, of course, even if all the above questions are set aside, how likely is it that the inferred result is actually correct, because there are still risks that extraordinary meditative experiences are, after all, not similar to a categorical basis. In addition, if there is more than one inferred result and they are in conflict with each other, it is not obvious that we can decide which one
is more likely. In the worst case, we might be unjustified to believe in any inferred result because of its lack of likelihood.

I shall not discuss issues surrounding the relation between CA2 and meditation in any further detail. This is because RM is, after all, a doctrine I will argue against. In addition, I, as a non-practitioner and non-researcher of meditation, am not qualified to discuss any more about meditative experiences. What I am arguing here is merely that if RM is true, CA2 has a chance of success, even though it is difficult to determine how much the chance really is.

9.5. The paradox of phenomenal judgment

In Section 9.4, I argued that CA has a chance of success if RM is true. However, in this sub-section I will put forward an argument to show that RM, with the CA or not, has a paradoxical epistemology of qualia, which does not allow for any belief or memory about qualia, and hence both CA and RM are implausible.

To tell the story of the paradox, we must begin with one major version of dualism: epiphenomenalism, according to which physical states cause qualitative states but not vice versa. It is commonly considered to have a self-defeating nature that cannot be easily solved, which presumably has a significant contribution to its current unpopularity. The worry is as follows. If qualitative states cannot cause anything physical, it seems that our judgments with regard to qualia – or, in Chalmers’s words, phenomenal judgments (1996, p. 177) – which our physical cognitive system has and can express to others by physical means, cannot be explained by qualia but solely by the physical. This is self-defeating. Chalmers calls the problem the paradox of phenomenal judgment (p. 177), and David Braddon-Mitchell and Frank Jackson call it the epiphenomenal objection to epiphenomenalism (2007, p. 13). (Henceforth, I will call it PPJ.)

Several authors, including John Hawthorne (2001, pp. 371-372), J. J. C. Smart (2004, p. 48), and David Braddon-Mitchell and Frank Jackson (2007, p. 141), use the same argument against RM, simply replacing the epiphenomenalist’s qualititative states with the Russelian Monist’s qualia-realising categorical properties. But Chalmers (1996) and William Seager (2009), in response to the worry, argue that RM is not a version of epiphenomenalism, and I will explain why I do not find their arguments objectionable. Nevertheless, I will provide an original argument to show that regardless of whether it is a version of epiphenomenalism, it is facing PPJ just as epiphenomenalism is. While many influential writers, such as Chalmers (1996), Daniel Stoljar (2014), John Heil (2004a), and
Seager (2009), consider it to be a significant improvement over dualism or epiphenomenalism, PPJ shows that the improvement might be overrated.

9.5.1. Epiphenomenalism and PPJ

Epiphenomenalism is the view that mental states are caused by physical states but have no effects upon physical states. In contemporary philosophy of mind, one of the most influential versions of epiphenomenalism is the version proposed by Frank Jackson (1982) and currently defended by Jaegwon Kim (2005). In this view, qualia are non-physical entities, and while they are causally affected by brain events, they are ‘totally causally inefficacious’ (Jackson 1982, p. 133).

But then the causal inefficacy of qualia leads to PPJ. According to the paradox, if epiphenomenalism is true, our phenomenal judgments, which require our physical cognitive system to possess, cannot be responses to and be explained by qualia. In particular, our phenomenal judgments include epiphenomenalism itself because it is a theory of qualia. Instead, our phenomenal judgments in them are all fully and exclusively accountable by physical facts. As Daniel Dennett notes, our zombie replicas would have made exactly the same judgments as us (1991, p. 403). Hence, as Dennett, and Braddon-Mitchell and Jackson argue, our phenomenal judgments will all become unjustified (Dennett 1991, p. 403; Braddon-Mitchell & Jackson 2007, p. 142). Braddon-Mitchell and Jackson even go further and argue that the case inevitably leads the epiphenomenalist to convert to physicalism, because within her metaphysical framework, her reason to posit anything beyond the physical must be fully and exclusively accountable by the physical, even though she may not actually know how (pp. 141-142).

Several critics of RM, including John Hawthrone (2001, pp. 371-372), J. J. C. Smart (2004, p. 48), and David Braddon-Mitchell and Frank Jackson (2007, p. 141), use the same argument against the doctrine, simply replacing the epiphenomenalist’s qualitative states with the Russelian Monist’s qualia-realising categorical properties. However, while the combination problem with RM, according to which it is difficult to see how categorical properties can constitute qualia, has drawn a lot of attention of contributors to the discussion of RM (see Section 11; see also the recent collection of

---

44 Jackson no longer believes in the view.

45 What is unique about this version of epiphenomenalism is that it takes only qualia, not every mental state, to be non-physical and is thus consistent with the standard view in neuroscience that non-qualitative and functional states of the brain are physical. This is different from older versions of epiphenomenalism that consider every mental state to be non-physical (see Campbell 1984 for the distinction between old and new epiphenomenalism).

46 Here I consider only concrete properties and entities, and set abstract properties and entities aside.
articles on RM edited by Torin Alter and Yujin Nagasawa, *Consciousness in the physical world: perspectives on Russellian Monism* [2015]), the relation between the doctrine and PPJ have not been discussed by its critics and defenders in much detail. Even in the cases of the authors I listed above, since the relation between RM and PPJ – and even RM itself – are not the central matters of concern of their work, their discussions are all quite brief (at most, a few paragraphs), and the perspective and possible responses of the Russellian Monist have not been considered in depth and with sympathy. What I wish to provide and develop is a more thorough and comprehensive inquiry into the relation.

**9.5.2. RM versus epiphenomenalism**

RM has recently attracted attention in the literature. It is considered by many influential writers to be an improvement over dualism or epiphenomenalism. Stoljar, for example, believes that it is a view that can preserve the ‘spirit and structure’ of standard physicalism and ‘so avoids the inelegance of dualism’ (2014, p. 19). Heil believes that RM abandons the mental-material distinction of dualism without falling into functionalism, which cannot take qualia seriously (2004a, pp. 239-240). Seager believes that RM is one of the most prominent alternatives to dualism, the ‘deeply unsatisfactory’ doctrine that commits to a ‘totally mysterious causal interaction between completely disparate substances’ (2009, p. 208). Chalmers, seeing PPJ as a problem to be avoided because of its counterintuitiveness (despite the fact that he finds it uncompelling because of the reason I will discuss in Section 9.5.5.1), particularly considers RM to be one of the ‘strategies for avoiding epiphenomenalism’ (1996, p. 151) and, in addition, considers it to be ‘the strategy to which [he is] most drawn’ (1996, p. 153). While I will express my agreement with Chalmers’s and Seager’s arguments that RM is a version of epiphenomenalism, I will provide an original argument that shows that the doctrine is still facing the same problem, PPJ, and hence the improvement might be overrated.

Before I provide the original argument, we shall look closely at why some believe that RM is not a version of epiphenomenalism and is thus not facing PPJ or similar problems, even though categorical properties are non-dispositional and non-causal much like epiphenomenal qualia are. In Chalmers’s view, there is a causal relevance between physical causal properties and their categorical bases, even though it would be, in his words, ‘a subtler kind of causal relevance than the usual kind’ (1996, p. 154). In his view, a world of physical causal and dispositional properties without categorical bases is ‘arguably logically impossible’ (p. 154) because it is not very reasonable to think that properties such as these are ‘empty placeholders with no properties of their own’ (p. 153). Some readers might be reminded of categoricalism, which we have discussed in Sections 6 and 7, according
to which causal and dispositional properties always have categorical bases. Indeed, I believe that Chalmers is simply restating the categoricalist view that causal and dispositional properties require categorical bases and the *empty space argument* (as I call it) in the literature, according to which dispositions without categorical bases to ‘fill in’ their spatial locations are nothing more than empty spaces and are thus impossible (see, for example, Armstrong 1961; Blackburn 1990). In sum, categorical bases, in his view, are not epiphenomenal because they are acting as necessary bases for physical causation to instantiate and are thus not metaphysically irrelevant to it. Let us call this argument the *argument from instantiation*.

Seager proposes another argument, which particularly relies on weak categoricalism, according to which dispositional properties are determined by the essences of (or are identical to) categorical properties (see Section 6.2). He writes:

> The fundamental dispositional properties of matter are just a reflection of its mentalistic intrinsic nature; calling them physical with the implication that they are entirely non-mental comes close to begging the question. (Seager 2009, pp. 217-218)

Seager is a panpsychist (not a non-standard physicalist like Stoljar or a neutral monist like Russell), but I do not think that his argument is dependent on panpsychism. In my understanding, his view can be paraphrased as follows. Dispositional properties are dependent on the nature or essence of their categorical bases, and in addition, the two kinds of properties are of the same matter. Hence, the two kinds of properties are not fundamentally distinct. Therefore, physical and mental properties, which are identical to or supervene on the two kinds of properties respectively, are not fundamentally distinct like the case of epiphenomenalism. Let us call this argument the *argument from unity*.

The strategy of both the argument from instantiation and the argument from unity is to say that there is a dependent metaphysical relation between dispositional properties and that of categorical properties, and in particular, they are categorical properties on which dispositional properties are dependent. I do not find these arguments to be objectionable if their metaphysical assumptions (i.e. categoricalism or weak categoricalism) are true. This is because if epiphenomenalism is defined as the view that the kind of property of concern (such as mental properties or categorical properties) has no ‘influence’ (in any sense) on physical causal properties, not as the view that the kind of property of concern has no causal effect on physical causal properties, then categoricalism and RM, as the two authors argue, should not be categorised as versions of epiphenomenalism. But I believe this is a
matter of definition, which I do not consider to be the crucial matter of concern. What matters is whether RM is facing PPJ like epiphenomenalism is, not whether RM matches with the definition of epiphenomenalism. And I will show that, indeed, it is by providing an original argument.

9.5.3. The zombie basis argument

I call my argument the zombie basis argument. It is, much like Chalmers’s zombie argument, an argument from conceivability. However, unlike the zombie argument, which invites us to conceive of philosophical zombies (i.e. perfect physical duplicates of us who lack qualia), the concept of physical duplicates is not involved in my argument. What I invite the reader to conceive of are (1) a kind of categorical property that is without qualia-realising nature, i.e. is not qualitative and cannot constitute anything qualitative and (2) a world containing it. With this in mind, sentient beings that have these categorical properties as their categorical bases are necessarily lacking in qualia (much like Chalmersian zombies are). Let us call these categorical properties zombie bases.

The zombie basis argument invites us to conceive of the following possible world. Possible world \( z \) is structurally and dynamically identical to our actual world, in the sense that it shares all the causal, dispositional, and structural properties. However, all of these properties are borne by zombie bases instead of their actual ones. With this in mind, our counterparts in world \( z \) would all have phenomenal judgments (and, of course, judgments with regard to all other matters) exactly the same as ours. Nevertheless, they still do not have qualia because they are borne by zombie bases.

If world \( z \) is really possible, whether our categorical bases are zombie bases or not makes no difference to our phenomenal judgments. Our phenomenal judgments are thus fully accountable by non-categorical properties and the bearing of these properties by any kind of categorical basis, which world \( z \) contains. Hence, PPJ persists. In addition, since the metaphysical and/or explanatory relevance between our non-categorical properties (such as causal, dispositional, and structural properties) and their categorical bases – as Chalmers’s argument from instantiation and Seager’s argument from unity point out – is the bearing relation we can find in world \( z \), it offers us no help at all in solving PPJ.

---

47 While I assume that most readers are familiar with Chalmers’s zombie argument and thus will not introduce it here, those who are unfamiliar with it can return to Section 9.3.3 for an introduction.

48 Our actual categorical bases are assumed not to be zombie bases for the sake of the argument, because Russellian Monism is now our default position to be objected. However, without the assumption that Russellian Monism is true, it is possible that our actual categorical bases are more likely to be zombie bases.
The zombie basis argument, if correct, shows that the bearing relations between dispositional properties and their categorical bases that Chalmers and Seager posit are insufficient to account for our phenomenal judgments about qualia, and hence there must be at least one mistake in their arguments. But it does not tell us explicitly what the mistake is. Nevertheless, I believe that it gives us a good hint about how to locate the mistake, which I believe to be the confusion between the two questions below:

**The structural question:** What is the systematic explanation of the fact that our brain system can output information about our qualia?

**The tracking question:** How can our brain system output information about our qualia that tracks and represents, in the sense of appropriately relating to and accurately describing, them?⁴⁹

The bearing of dispositions by their categorical bases, as Chalmers and Seager point out, contributes to our answer to the structural question, in exactly the same way that it contributes to our inquiries into the structure of all material objects, functions, and states of affairs – which include a chair, my Microsoft Word, the eliminativist’s reason for believing that there are no qualia, and so forth. Unfortunately, what the zombie basis argument shows us is that this alone cannot be a tracking relation that can help us to answer the tracking question. Worse, it is difficult to see how there can be such a relation within the theoretical framework of RM because whatever information our brain outputs, and all the structural and dynamical features it metaphysically supervenes on, are shared by its counterpart in world z, the world of zombie bases.

With this in mind, we can conclude that Chalmers and Seager either have the confusion that bearing alone is sufficient to explain informational tracking, or have missed the point that what is really in need of explanation is informational tracking. Either way, their arguments cannot offer RM any help in solving PPJ.

### 9.5.4. Implications of the argument

I believe that the implications of the zombie basis argument are simple and straightforward. This is because the purpose of the argument is simply to show that PPJ persists in RM. With this in mind, if it is correct, all of the implications of PPJ to epiphenomenalism apply to RM as well. In fact, when

---

⁴⁹ I am heavy indebted to Frank Jackson for a discussion with respect to this point.
Braddon-Mitchell and Jackson argue that PPJ inevitably leads the epiphenomenalist to convert to physicalism because any of her reasons to posit something beyond the physical must be fully and exclusively accountable by the physical, they also have RM in mind (2007, pp. 140-141). The difference between their argument and my argument here is that they simply assume that RM is a version of epiphenomenalism, while I avoid this assumption and critically assess Chalmers’s and Seager’s arguments that RM is not a version of epiphenomenalism. What I have (hopefully) shown is that even if Chalmers’s and Seager’s arguments that RM is not a version of epiphenomenalism are correct, PPJ persists.

With this in mind, the improvement of RM over dualism (and epiphenomenalism in particular) might be overrated. It is, contrary to many beliefs, as untenable as they are. Even if the categoricalist is right that every dispositional property has a categorical basis, we have no reason to believe that the categorical bases are qualitative, or can constitute qualia in some manner that is not (solely) via the causal, dispositional, and structural properties that are, in turn, borne by these bases. Returning to the main purpose of this section, this implies that categorical acquaintance thesis (CA) is implausible if it is grounded on the metaphysical framework provided by RM. In addition, even if RM is true, since the existence of qualia will then have no influence on our phenomenal judgments, CA still must be false.

9.5.5. Possible objections

9.5.5.1. The so what objection

The so what objection (as I call it) is offered by Chalmers (1996). It does not target the zombie basis argument but PPJ, which he believes to be not compelling.50 According to the objection, there may indeed be a counterintuitive disconnection between epiphenomenal qualia and our judgments about them, but he asks, ‘So what?’ (p. 198). In his view, there is a difference between phenomenal judgments and phenomenal beliefs: our zombie replicas share our phenomenal judgments, but they do not share our phenomenal beliefs precisely because of the qualitative experiential element of the beliefs they lack (pp. 207-208). With this in mind, a lack of functional difference does not imply a lack of experiential difference, and our immediate qualitative experience justifies our phenomenal

50 Chalmers does not consider PPJ to be compelling (1996, p. 160). However, with regard to the view that PPJ ‘is a fatal flaw’, he claims that he has ‘some sympathy with the position’ (p. 160). This is why he provides ways we can avoid epiphenomenalism, which include Russellian Monism and the argument from instantiation used to defend it.
judgment and belief that non-physical qualia exist. He writes, ‘I know I am conscious, and the knowledge is based solely on my immediate experience’ (p. 198, original emphasis); ‘It is having the experiences that justifies the beliefs’ (p. 196, original emphasis).

My reply

Let us set the Russellian Monist aside and focus on the epiphenomenalist for the moment. We can understand the point of view of the epiphenomenalist with more sympathy and agree that, from her point of view, for something to be called a belief about qualia, it must be two components, which are the judgmental component and the qualitative component. The judgmental component is fully and exclusively accountable in terms of physical causation and, in addition, causally efficacious to other physical entities and the qualitative component, while the qualitative component is non-physical, causally impotent, and immediately acquainted. This is different from most (standard) physicalists who believe that the judgmental component alone is sufficient to be called a belief. However, even if we assume this point of view, I do not believe that this leads us to Chalmers’s conclusion.

What actually follows is that the judgmental component has a radical disconnection from and independence of the qualitative component. The judgmental component serves as the ground of a belief about qualia, while the qualitative component serves as, in J. J. C. Smart’s famous phrase, ‘nomological danglers’ (1959, p. 142) passively determined by but not metaphysically supervening on the judgmental component. As a result, the qualitative component could not add anything new to the story of why we make our judgments, which is only about the physical. It cannot get into the picture of the physical and help us to explain why our beliefs about qualia, including Chalmers’s judgment that he is having some immediate belief that is different from his zombie replica, are brought to existence and are justified.

With this in mind, it seems that Braddon-Mitchell and Jackson’s comment on the epiphenomenalist, which we have discussed earlier, is correct. If she wants the qualitative component to be a part of the story about how we come to have our beliefs about and theories of qualia, she will be forced to expect that it is within the realm of physical causation like the judgmental component is. Unfortunately, this move inevitably converts her to physicalism, since qualia should now be fully and exclusively accountable in terms of the physical, even though we might not yet know the account. And if the zombie basis argument is right that the problem of epiphenomenalism persists in the case of RM, what Braddon-Mitchell and Jackson’s comment on the epiphenomenalist presumably applies to the Russellian Monist as well.
No doubt, Chalmers can continue to stress his confidence in his possession of qualia. In fact, I am sympathetic to this claim. Nevertheless, I believe that it is one thing to be confident that one possesses qualia, and another to be confident of a theory of qualia that is facing PPJ. What I believe one should do in the face of PPJ and her confidence in her possession of qualia is to avoid accepting a theory of qualia that is facing PPJ, not to take PPJ as un compelling. Chalmers argues that alternative views (which presumably include physicalism and interactionist dualism) ‘are simply wrong, as we have seen again and again’ (1996, p. 160). However, not everyone – myself included – shares his intellectual journey. Furthermore, even if we accept his perspective, perhaps we simply ought to be agnostic and accept the fact there are no good theories of consciousness yet, not that those facing PPJ are the best or the least problematic. Hence, I find Chalmers’s view that his confidence of his possession of qualia could resist the compelling force of PPJ a bit puzzling.

9.5.5.2. The objection from metaphysical necessity

According to some theories of categorical properties, there is a metaphysically necessary relation between dispositional properties and their categorical bases, which do not allow a dispositional property to be multiply realisable by different categorical bases (see, for example, Heil 2005). In this case, world z and zombie bases bearing our dispositional properties would be inconceivable and impossible. For example, in response to Chalmers’s zombie argument, Heil writes, ‘Agents or systems possessing identical powers must be qualitatively identical as well’ (2003, p. 247).

My reply

It is not easy, at least for me, to see why our dispositional properties cannot be borne by categorical properties with different or no qualia-realising nature. Even Heil, who argues that categorical properties are identical with dispositional properties because they are their ‘qualities’, still somehow describes dispositional properties and their ‘qualities’ as if they are taking turns to manifest in different situations – such as scientific measurements and our qualitative experiences (see Heil 2003). This seems to imply that those dispositional properties and qualities can be combined in some different ways, at least in some distant possible worlds very much unlike our actual world (even though he himself may not agree).

51 Heil’s identity theory does not allow for multiple realisability and is thus much stronger than the identity theory used by Smart (2007) to explain the mind/body relation.
Perhaps there is a metaphysically necessary relation between the *qualia-realising nature* and the *disposition-realising nature* of a categorical property. In this case, the metaphysically necessary relation is either (1) brute and inexplicable or (2) explainable by some further fact. (1) In the former case, it is certainly true that a brute and inexplicable metaphysically necessary relation easily blocks the metaphysical possibility of world $z$ and zombie bases bearing our dispositional properties. However, the use of a brute and inexplicable metaphysically necessary relation to argue against the zombie basis argument is a bit puzzling. First of all, the proposal that brute and inexplicable metaphysically necessary relations exist is itself puzzling and controversial. This is because it is an easy – rather too easy – response to so many philosophical arguments, while the nature of any such relation by definition cannot be precisely described. Hence, the worry is that the existence of such things is simply a posit for an *ad hoc* purpose, and thus we have no reason to believe in it.

Secondly, even if the metaphysical necessity is brute and inexplicable, it seems that the absence of such a metaphysical necessity is conceivable. With this in mind, regardless of whether world $z$ and zombie bases bearing our dispositional properties are metaphysically possible, they are conceivable. In this case, conceivable world $z$ can still be considered a theoretical model. Compare it with the theoretical description of the actual world by the Russelian Monist prior to the addition of the brute and inexplicable metaphysically necessary relation. In terms of their content, they are not only structurally and dynamically alike, but also cannot have any difference in their descriptions of the epistemic structures of their habitants. Hence, conceivable world $z$ seems to sufficiently show us that the Russelian Monist, prior to the addition of the metaphysically necessary relation, has not provided any tracking relation that can solve PPJ *in her theory*. No doubt, now the objector could add the brute and inexplicable metaphysically necessary relation into the world described by the Russelian Monist, but can it serve as an informational tracking relation? It is very difficult to see how it could be so. For me, this is, in fact, fairly unintelligible.

On the other hand, if it is expected that the metaphysically necessary relation between the qualia-realising nature and the disposition-realising nature (2) can be explained by some further fact, then I believe it is either (i) unlikely to be true or (ii) likely to be damaging to RM. This is because there will be an *explanatory gap* between the disposition-realising nature and the qualia-realising nature that is very similar to the explanatory gap faced by the (standard) physicalist. The idea of the explanatory gap, as it is influentially introduced by Joseph Levine (1983), is as follows. While (standard) physical properties are causal in nature, qualia such as pain and itch are non-causal in nature. With this in mind, the causal stories we can have by taking (standard) physical properties into

184
account cannot tell why qualia feel the way they do. Even though Levine merely intends to point out that the explanatory gap leaves us inevitably agnostic about which particular mind-body theory is true (p. 359), many take the further step to reach the metaphysical conclusion that (standard) physicalism is false. For example, Chalmers develops the explanatory gap into a principle which Stoljar calls from-structure-only-structure, according to which ‘from structure and dynamics, we can only get more structure and dynamics’ (Chalmers 1996, p. 118; Stoljar 2006, p. 144, 2014, p. 29, 2015a, p. 325); and according to Chalmers, all of the major arguments against (standard) physicalism, including his zombie argument and the Mary’s room argument, start by establishing it (2003a, p. 107).

I believe that the same gap applies to the metaphysically necessary relation between the disposition-realising nature and the qualia-realising nature of a categorical property. If it is considered to be explainable by some further fact instead of being brute and inexplicable, it is difficult to see what kind of fact can connect disposition-related facts and qualia-related facts. Returning to my point that it is either (i) unlikely to be true or (ii) likely to be damaging to RM. The reason is as follows: (i) On the one hand, it is difficult to see how the gap can possibly be bridged. (ii) On the other hand, even if the gap can be bridged in some ways I have never considered, it is probably bad news rather than good news for the Russellian Monist. That is because it is likely that we can reject Chalmers’s from-structure-only-structure principle by having a very similar (standard) physicalist theory that posits a similar explanatory relation between causal or dispositional properties and qualia. The Russellian Monist will then face the risk that she might lose her motivation to get beyond the (standardly) physical and appeal to RM to explain qualia in in the first place.

In sum, appealing to metaphysical necessity to resist the possibility of world z will lead one to fall into a trilemma. (1) If the necessity is brute and inexplicable, then it cannot really reject the conclusion of the zombie basis argument. (2i) If the necessity is considered to be grounded on some further fact, then it is difficult to see how there can possibly be such a fact. (2ii) Even if there can possibly be such a fact, it is very likely that such a fact may undercut our motivation to get beyond (standard) physicalism to believe in RM.

9.5.5.3. The weak conceivability objection

The weak conceivability objection specifically targets the zombie basis argument. The objection can be developed from Stoljar’s work, even though Stoljar does not have the zombie basis argument in mind (see Stoljar 2001b; see also my introduction to his argument in Section 8.3.3). What he has in mind is, instead, the worry about the possibility of zombies that do not possess only our standard
physical properties but our categorical bases as well. In response to the worry, he argues that we cannot ‘truthfully be said to strongly conceive the possibility of something which is [categorically] identical’ (2001b, p. 401). This is because our ignorance of categorical properties does not allow us to strongly conceive of the possibilities concerning them. A similar argument can be made against the zombie basis argument. It can be said that zombie bases bearing our standard physical properties are not truthfully and strongly conceivable because of our ignorance of categorical properties.

My reply

A problem with this objection is that it might be too strong. We rarely think about possibilities and necessities by taking a full and comprehensive set of knowledge of the entities we are considering into account. As Chalmers notes, it is ideal conceivability that entails possibility (2002, p. 204), but ‘it is not obvious that an ideal reasoner is possible or coherent’ (p. 148). If we were to accept the fact that our (practically possible) conceivability does not imply possibility, then there would be a disastrous overkill. We may, suddenly, lose most of our strong arguments against (standard) physicalism and thus have no reason to find RM attractive.

The reason is as follows. As Chalmers notes, all major arguments against (standard) physicalism begin with the explanatory gap (2003a, p. 107); and the explanatory gap, as I understand it, rests on the inconceivability and thus impossibility that features of the (standardly) physical and qualia can be bridged. If conceivability does not imply possibility, then it seems that we should just agree with Patricia Churchland (2002) that the gap is merely a matter of current ignorance and nothing metaphysical, and that given the recent progress of neuroscience, it is difficult to see why we should not expect that neuroscience will eventually provide the solution.

Chalmers, I believe, convincingly provides a proposal that can allow us to have – or, in my view, have something closer to – ideal conceivability. The proposal is that we may have a justification that a belief is undefeatable by better reasoning. As he puts it:

[O]ne can […] simply invoke the notion of undefeatability be better reasoning. […] We can say that S is ideally conceivable when there is a possible subject for whom S is prima facie conceivable, with justification that is undefeatable by better reasoning. The idea is that when prima facie conceivability falls short of ideal conceivability, then the claim that the relevant tests are passed will either be unjustified, or the justification will be defeatable by further
reasoning. For ideal conceivability, one needs justification that cannot be rationally defeated. (2002, p. 148)

This line of reasoning is what I believe he has (or tacitly has) in his earlier work, in which he writes:

[W]e do not need to imagine each of the neurons to make the case. Mere complexity among neurons could not conceptually entail consciousness; if all that neural structure is to be relevant to consciousness, it must be relevant in virtue of some higher-level properties that it enables. So it is enough to imagine the system at a coarse level. (1996, p. 98)

We can see that Chalmers obviously has the from-structure-only-structure principle we discussed earlier. I believe that his line of reasoning is as follows. Our currently unbridgeable explanatory gap implies something in principle, which is the from-structure-only-structure principle. If something is in principle, then it is undefeatable. This, in turn, implies ideal conceivability of zombies, which, in turn, implies the possibility of zombies. The crucial step is to discover something in principle.

Let us return to the conceivability of world z and zombie bases bearing our dispositional properties. With Chalmers’s line of reasoning, to guarantee their possibility, we do not need to have perfect knowledge about categorical properties; we only need to have something in principle with regard to their nature. Of course, the matter of concern, as Stoljar points out, is whether the categorical ignorance thesis (CI) allows us to discover it. But I think the answer is obviously yes. CI merely prevents us from knowing what categorical bases are precisely like (see Section 7.2). On the other hand, metaphysical issues such as realisation relation, identity relation, and so forth are more abstract and imprecise. We can still assess which kinds are plausible, in terms of whether they make sense in logic and metaphysics, whether they are compatible with natural science, and so on. This allows us to have some metaphysical principles with regard to the nature of categorical properties. For example, in Section 9.5.5.2, based on some general metaphysical principles, I spelled out the possible scenarios in which zombie bases are metaphysically impossible. It is difficult to see why CI would be against it.

Now it should be clear how we can respond to the worry that zombie bases are not strongly conceivable. In Section 9.5.5.2, I argued that the possible scenarios in which zombie bases are metaphysically impossible force the Russelian Monist into a trilemma. Since this is based on some
general metaphysical principles, and CI is not helpful in avoiding something in principle, it cannot help us to avoid the trilemma.

9.6. Conclusion

In this section, with the assumption that all objections against RM have no force, I critically assessed several possible ways the categorical acquaintance thesis (CA) can work. In particular, I argued that using inference to the best explanation in combination with meditative experience has a chance of success, even though it is difficult to assess how much the chance really is. However, by the zombie basis argument, I then argued that RM is facing the paradox of phenomenal judgement (PPJ) just as epiphenomenalism is. If this is correct, RM has a paradoxical epistemology of qualia, which does not allow for any belief or memory about qualia. Hence, both the doctrine and CA are implausible and thus have no threat – potential or actual – to both metaphysical naturalism and the categorical ignorance thesis (CI).
Part III

Section 10. The naturalistic unacceptability of Russellian Monism

10.1. Introduction

In Section 8, I sketched Daniel Stoljar’s Russellian non-standard physicalism, according to which posits of Russellian Monism (RM) can be categorised as physical. In this section, I will answer the question: can RM can be a version of or a doctrine similar to (physicalism and hence) metaphysical naturalism (MN)? Recall that physicalism is a version of MN, and hence physicalistic respectability implies naturalistic respectability. Recall also that the importance of the questions is as follows: (I) If RM can be a version of MN, then the categorical acquaintance thesis can be grounded on naturalistically respectable assumptions, which may allow it to resist the possible objection that it is not naturalistically respectable. (II) If RM cannot be a version of MN, it may still be an attractive doctrine and be similar enough to MN, which allows it to avoid most shortcomings of other versions of metaphysical non-naturalism and thus be the most formidable rival of MN.

In Section 10, I provided an argument that shows that RM is untenable. If the argument is correct, all challenges the doctrine can possibly provide to MN and the categorical ignorance thesis (CI) can be said to be without force. However, for the interest of some readers, it might be useful to nevertheless go on to assess whether philosophers like Stoljar are correct in claiming that RM can be a version of or a doctrine similar to (physicalism and hence) MN. In this section, I will focus on Stoljar’s Russellian non-standard physicalism – arguably the most well-developed attempt to develop RM into a version of physicalism – and argue that, contrary to Stoljar’s view, RM must ultimately collapse into dualism and thus cannot be a version of physicalism or MN. In addition, it cannot even offer us a description of the world that is similar to one of MN or physicalism.

What I will argue is roughly the following. Let us assume that Stoljar’s way to define the physical and physicalism is correct. While it may appear that RM is consistent with (or almost consistent with) the definition, it is actually not the case. This is because what contents the Russellian Monist believes her theory has is one thing, and what contents the theory must have in order to be plausible is another. I will argue that any plausible version of RM cannot be consistent with Stoljar’s conception of physicalism, nor can it provide a description of the world that is similar to Stoljar’s physicalism. And
I will provide additional arguments to show that for the same reason this can also be said about the relation between the doctrine and the version of MN I provided in Part I.

In particular, what forces RM to be physicalistically unacceptable (in the sense Stoljar defines physicalism) and naturalistically unacceptable (in the sense I define MN in Part I) is the so-called combination problem in the literature, which is (roughly) the problem of how our categorical bases can ‘combine’ into and constitute our qualia. More details will be provided as the discussion goes on.

10.2. The problem facing Stoljar’s Russellian non-standard physicalism

Recall that in Stoljar’s view categorical properties can be defined as physical, despite the fact that they are not posited by physical theories, nor are they constituted by the posits of the theories. This is because of his object-based conception of the physical, according to which the physical can be defined with reference to properties required to account for paradigmatic physical objects and those that supervene on these properties. And by the term ‘paradigmatic physical objects’, I have argued that he seems to mean objects we intuitively consider to be physical or objects most of us agree to be physical (or something very similar) (see Section 8.3.2).

The view, I believe, is a very attractive one because of his argument from cost (as I call it). Recall that the argument is as follows. If we accept only the theory-based conception of the physical, then categorical properties cannot be counted as physical. Physicalism would be false because of metaphysical assumptions about dispositions that have nothing to do with philosophy of mind (2001a, p. 272). In a personal correspondence, he also added that David Lewis, David Armstrong, and many others would fail to be physicalists (p.c.). No doubt, this is inconclusive from the perspective of this dissertation because there are some who are willing to accept (or lean toward) the view that categorical properties are essentially non-physical (Langton 1998; Braddon-Mitchell & Jackson 2007; Foster 1993; Ney 2007), and it is exactly one of the purposes this dissertation to assess the view. Nonetheless, this will be left to be discussed in detail in Part IV, and will be set aside at this moment. For the sake of argument, we can assume that Stoljar is correct.

It is tempting to agree with Stoljar that the Russellian Monist should be considered to be a physicalist because what the Russellian Monist adds to our picture of the world is exactly like what many respected physicalists such as Lewis, Armstrong, and many others do, which are unknowable categorical properties. The disagreement between the Russellian Monist and, say, Armstrong is only about whether a component of RM, the categorical qualia thesis (CQ), is true. Recall CQ:
CQ: Our qualitative experiences (i.e. qualia) are (solely or partially) constituted by their categorical bases (in some manner that is not [solely] via the causal, dispositional, and structural properties that are, in turn, borne by these bases).

Despite the existence of such difference, most physicalists consider entities or properties that are solely constituted by (or supervene on) other physical properties (such as a tree or the Statue of Freedom) to be physical. Since what the Russelian Monist believes to be qualia are constituted by categorical properties that we should take to be physical because of the views of Lewis, Armstrong, and so on, we should also count such qualia as physical.

However, I believe that there is a further worry about Stoljar’s view even if all of the arguments above are correct. The worry is that what he proposes is still a very rough and basic metaphysical framework, which tells us little about how the mind actually operates and is structured. But RM is a theory in philosophy of mind, in which such details have to be provided; and in the process of filling in such details, it is possible that we will inevitably commit to non-physicalism, even in the sense Stoljar defines it. Stoljar notices the need to fill in details. He admits that his purpose is only to see Russelian non-standard physicalism as ‘a kind of epistemic ideal’, which precisely is a prediction that inquiries into categorical properties are needed to understand qualia and their place in the world (2001a, p. 274). In addition, he ‘agree[s] this is a drawback’ that he ‘offer[s] no constructive suggestions about how that goal is to be reached’ (2001a, p. 275). Nonetheless, I will argue that it is not only a drawback caused by a missing piece, but a fatal problem facing his view if we carefully assess what kind of hypothesis can actually be the piece that fills in the gap.

10.3. The combination problem

The particular problem I will focus on is the combination problem. The problem is simple and straightforward, which is how our categorical bases can ‘combine’ into and constitute our qualia. The term ‘combination problem’ is firstly introduced to the literature by William Seager in his article, ‘Consciousness, information, and panpsychism’ (1995), in which he uses the term to name an argument against panpsychism seminally made by William James in his Principles of psychology (1918/1890). In James’s view, panpsychism is ‘theoretically […] very confused’ and ‘inwardly quite unintelligible’ (p. 145). One argument he proposes against the doctrine is that multiple consciousness
in principle does not combine. Two independent consciousness in principle do not sum up and combine into a third consciousness. James’s following passage is still frequently cited by contemporary writers:

Where the elemental units are supposed to be feelings, the case is in no wise altered. Take a hundred of them, shuffle them and pack them as close together as you can (whatever that may mean); still each remains the same feeling it always was, shut in its own skin, windowless, ignorant of what the other feelings are and mean. There would be a hundred-and-first feeling there, if, when a group or series of such feelings were set up, a consciousness belonging to the group as such should emerge. And this 101st feeling would be a totally new fact; the 100 original feelings might, by a curious physical law, be a signal for its creation, when they came together; but they would have no substantial identity with it, nor it with them, and one could never deduce the one from the others, or (in any intelligible sense) say that they evolved it. (p. 160)

Most writers on RM take the problem to be the most important problem facing the Russellian Monist (even though I disagree because I believe that the paradox of phenomenal judgement [PPJ] I discussed earlier is a much more significant problem); and the problem has independently and dependently arisen in the writings of a number of writers commenting on RM and panpsychism. Examples include Campbell (1984), Van Cleve (1990), Searle (2004), Hohwy (2005), Goff (2006, 2009), Macpherson (2006), Dainton (2011), Chalmers (2015, forthcoming), Bennett (manuscript) and many others. Even Chalmers, the most influential writer on RM and panpsychism, says that he is not confident that these doctrines are true because of the problem (2015, p. 247).

There can be many ways to inquire into the problem in a dissertation. One way is, of course, to carefully assess all possible solutions one by one to see if any can be successful. This way is, no doubt, comprehensive, but it cannot tell the exact problem that is common and crucial to all possible solutions, nor does it match our purpose, which is to see if RM can be physicalistically respectable or similar to physicalism. In addition, it is, certainly and unfortunately, impractical because of the limit of space. Hence, the way I will do it is, instead, to provide a dilemma for the Russellian non-standard physicalist, according to which there are two general ways to solve the combination problem that are available to her, but one way is implausible while the other is physicalistically unacceptable and
dissimilar to physicalism. This leads to the conclusion that all plausible versions of RM are physicalistically unacceptable and dissimilar to physicalism.

10.4. The dilemma of physical combination

I call the dilemma I will put forward the dilemma of physical combination. The dilemma is whether emergence is involved in the process of combination or not. I will argue that if the answer is no, any theory of mind we could develop is implausible; and if yes, any theory of mind we could develop is physicalistically unacceptable and dissimilar to physicalism. Therefore, any plausible and comprehensive version of RM must be physicalistically unacceptable and dissimilar to physicalism; the proposal of developing Russellian non-standard physicalism is not successful.

10.4.1. Combination without emergence

Let us begin with possible forms of combination without emergence. A combination as such involves the combined property or entity being nothing more than the sum of the lower-level properties or entities of concern and their relations (with the possible addition of properties or entities that are required to ground them, such as spatiotemporal properties). As Chalmers writes, categorical properties might ‘collectively constitute phenomenal properties when organized in the appropriate way’ (2010, p. 151); and as John Heil writes, qualia ‘owe their existence to the properties of the components of brains and their arrangement’ (2003, p. 235).

A question is, of course, what the relations or the arrangements of the categorical properties are. One possibility is spatiotemporal arrangements, or, in the case that spatiotemporality is not fundamental or not real, any arrangements with respect to whatever ideal physical science posits to replace it. Another possibility, as Jakob Hohwy, a critic of RM and proponent of the combination problem, reminds us, is the dispositional roles which categorical properties play (2005, pp. 84-85).

I believe that all possible proposals of combination without emergence are implausible because of what I call the problem of appropriate mereological sum. The problem is as follows. Consider the doctrine of unrestricted mereological composition in mereology, according to which any objects

---

52 No doubt, not every philosopher will agree with this rough conception. In particular, some proponents of non-reductive physicalism believe that some higher-level entities such as mind are not merely the combination of physical particles and their relations. However, I will set this aside because of two reasons. The first is that I believe the rough conception I have provided is sufficient for our purpose. The second is that I believe that the conception of non-reductive combination is a bit mysterious and unclear, and thus the burden of providing a definition of combination without emergence that is compatible with it should be left for its proponents.
whatever compose something. There can be nearly infinite mereological sums we can find in the world. While a rock is a mereological sum of its particles, we can also think of some less ordinary mereological sums, such as that of the moon and my computer, that of the fingers of Donald Trump and Vladimir Putin and my left hemisphere, and so on. The principle is, no doubt, highly controversial, but I am not assuming that it is true. The only concern that I am raising here is the amount and variety of possible objects there are if objects can be combined without restrictions. The problem of appropriate mereological sum then follows. Among the tremendously large number of possible composite objects, only a significantly small proportion of them possess qualia – or at least possess qualia like ours – in some non-excessive manner. (By a ‘non-excessive manner’, I mean the manner in which an object has a significant proportion of its parts responsible for the constitution of the qualitative experience. For example, my brain – or at least a part of my brain – not only possesses my qualia, but also has a significant proportion of its parts responsible for their constitution; on the contrary, the mereological sum of my brain and the Empire State Building, no doubt, possesses my qualia, but only a very insignificant proportion of its parts, which is my brain, is responsible for the constitution of my qualia.) Even the panpsychist would agree that while a rock possesses qualia, it does not possess qualia like ours (see, for example, Chalmers 1996, p. 299), because even we ourselves possess very different kinds of qualia when we are in different physical states, such as being fully awake, in a dream, and drunk. If we can all agree that the fact that only a small proportion of possible composite objects have qualia like ours in some non-excessive matter, what, then, are the conditions that explain the composition of those qualia? Certainly, we cannot have the condition unrestricted like in the way similar to the doctrine of unrestricted mereological composition, having any objects whatsoever to compose qualia like ours.

The very natural and only possible answer is the neurological function(s) of our brain. This answer looks attractive but is implausible, because neuroscience has taught us that our neurological functions are of a very high-level biological structure, which is, in turn, a very large-scale physical structure. The way these neurological functions operate and the way fundamental particles operate are very much indifferent to each other (see Smart 1978; Stenger 1995; Hohwy 2005); and it is a

53 No doubt, the mereological sum of my brain and a small object without qualitative experiences, such as a coin, would have a significant proportion of its parts constituting qualitative experiences. Nevertheless, the kind of mereological sum mentioned here is sufficient for our purpose. After all, not even our brain has every of its parts responsible for the constitution of qualitative experiences

54 For a different but similar and related problem, see Rosenberg 2015. Gregg Rosenberg puts forwards the boundary problem (as he calls it), according to which our consciousness has boundary and does not extend out of us, and it is not easy for the panpsychist to tell why (p. 227).
fundamental physical entity which a categorical property of concern of Russellian monism is identical to or bears the role of. As J. J. C. Smart describes, neurons are ‘ordinary matters’ (1978, p. 340), in the sense that the studies relevant to understanding them are the studies of cytological structure, chemistry, and electricity, not sub-atomic level matters. With respect to fundamental level matters, such as the motions of quarks, it is difficult to see how a brain is different from a rock – from the perspective of microphysics they are both a bunch of random quarks and nothing more – let alone the categorical bases of those quarks.

Another answer popular among panpsychist authors, though I consider it implausible, is that quantum physics provides us some clue in solving the combination problem (Seager 1995, p. 284; Lockwood 2015, p. 154; Rosenberg 2015, p. 228). As many other writers, such as Patricia Churchland and Chalmers, correctly note, the worry is that quantum mind theories lack any substantive empirical evidence in neuroscience (Churchland 2002, pp. 196-197; Chalmers forthcoming). With this in mind, recall Stenger’s argument above. Since the need for quantum physics to explain that our brain and a rock are not different, appealing to quantum physics, though popular among panpsychists, simply provides us no solution to the problem of appropriate mereological sum.

An answer to the combination problem that is popular among panpsychist writers is that, contrary to Smart’s ordinary matters thesis, some quantum mind theories are true (see, for example, Seager 1995, p. 284, forthcoming; Rosenberg 2015, p. 228). But it is noteworthy that we cannot simply reply to the problem of appropriate mereological sum by stating that our mind is constituted by quantum entities – of course it is, but so as a rock. There has to be a philosophical or scientific theory that shows that our mind involves quantum phenomena in some unique ways. In fact, there are indeed such theories. One influential example is physicist Roger Penrose’s theory of ‘orchestrated objective reduction’ (1989), according to which our brain is a quantum computer that makes use of some quantum effects in microtubules when it functions. However, for theories like that of Penrose to be taken seriously, there should be, firstly, empirical evidence in neuroscience provided; and as many writers, such as Patricia Churchland (2002, pp. 196-197) and Chalmers (forthcoming), correctly note, the worry is that such evidence is still lacking. In addition, as physicist Victor Stenger (1995) argues, even if theories like that of Penrose is correct, it is one thing to say that our brain is making

---

55 Seager and Lockwood in fact defend combination by emergence, advocating emergentist theories that take quantum physics into account. Nevertheless, I believe that a theory of combination without emergence that takes quantum physics into account is possible response to the problem of appropriate mereological sum, and is thus worthy of discussion.
use of some quantum effects in some of its neurological computational functions, and another to say that the function of our brain as a whole has to be understood as a single quantum phenomenon. Only the latter is helpful to answering the problem of appropriate mereological sum, because, as neuroscience tells us, our qualitative experiences are based on large-scale neural networks that spread across the brain, not any small scale quantum phenomenon within, say, a particular microtubule. However, the latter is also very inconsistent with the biology of the brain that neuroscience tells us, according to which a mental state is based on a collection of countless separate and independent neuron firings, not any single quantum phenomenon. This is, in my view, a sufficient reason to consider the response to the problem of appropriate mereological sum to be an implausible one.

10.4.2. Combination by emergence

Let us move to the second horn of the dilemma of physical combination, combination by emergence. Appealing to emergence can provide a solution to the problem of appropriate mereological sum in an easy way – rather too easy, but physicalistically unacceptable and dissimilar to physicalism (even in the way that Stoljar defines physicalism). According to emergentism of qualia, whenever there is a physical structure like our brain, a ‘novel’ property or entity which is not a sum of lower-level properties or entities (and their relations) will be brought into existence in some way inexplicable by the nature of the latter. And emergentism can be compatible with RM (or panpsychism) in the sense that the lower-level properties or entities that are required to act as the emergence bases include categorical properties (or, in the case of panpsychism, micro-qualia).

Philosophers who take emergence as a respectable solution to the combination problem include Keith Campbell (1984), James Van Cleve (1990), William Seager (2010, forthcoming), and Gregg Rosenberg (2015). For example, as Van Cleve puts it in his seminal article ‘Mind – dust or magic? panpsychism versus emergence’ (1990):

Let us ask, with what sort of mentality does our panpsychist wish to endow the fundamental particles? Presumably not with intelligence and perception, probably not even with consciousness; Nagel speaks merely of “proto-mentality.” Well, then, will intelligence in a complex system follow with logical necessity from the proto-mental properties of the constituents? I doubt it; the prospect of this seems no brighter than of the mental following from the purely physical. This means that we must admit a kind of emergence after all – of
the higher mental functions from the lower mental functions. So why not simply admit the possibility of emergence to begin with? (1990, p. 219)

The version of panpsychism provided by Van Cleve is not a version of RM because the features of consciousness he has in mind do not include only perception, but also intelligence, which is, presumably, a causal function. But his argument applies to RM as well if we set intelligence aside and focus on perception, especially its qualitative nature.

I shall not comment on the plausibility of emergentism about qualia because this is not our purpose here. Our purpose is, instead, to question whether the doctrine is physicalistically respectable or not. Unfortunately, the answer seems to be no. Since qualia in this view is novel, not the sum of their base properties, it is difficult to see how they can be understood as metaphysically supervening on their base properties. Hence, the view cannot be consistent with Stoljar’s object-based conception of the physical, according to which the physical can be defined with reference to properties required to account for paradigmatic physical objects and those that supervene on these properties. We simply do not need emergent qualia (and/or anything supervening on them) to account for paradigmatic physical objects like chairs, trees, and computers.

Some readers might not find my appeal to the novelty of emergent qualia compelling enough, because the relation between emergence and supervenience is a controversial issue. Many take the two to be compatible, or at least possibly compatible when adequately defined (see, for example, Van Cleve 1990). Here is a stronger supporting argument for my view. As David Braddon-Mitchell (2007) argues, for emergence to happen, only laws of nature that are responsible for the ordinary behaviours of the emergence bases are insufficient; and there must be some special laws of nature that are fundamental. This is because emergent properties are not merely the aggregate of the emergence bases. With this in mind, there would be always some possible worlds containing the aggregate of the emergence bases, but without emergence happening. Therefore, for our actual world to be differentiated from those worlds, there must be some additional laws of nature that are irreducible to ordinary laws and thus fundamental. In the case of qualia, they would be fundamental psychological laws. Are fundamental psychological laws or anything supervening on them required to account for paradigmatic physical objects like chairs, trees, and computers? The answer is, obviously, ‘No!’

An additional note. Some writers tend to believe that the emergence of qualia is not *sui generis*, but is a consequence or an extension of quantum emergence – which we may find some examples in current physical theories – within our brain (see, for example, Rosenberg 2015; Seager forthcoming).
This line of thought falls back to the kind of quantum mind theory I have discussed in Section 10.4.1; and I shall not repeat my response here.

Recall that I have argued that combination without emergence is implausible because of the problem of appropriate mereological sum. The dilemma of physical combination for the Russellian non-standard physicalist should now be clear. She may explain combination with or without emergence, but combination without emergence is implausible, and hence she, in fact, must make use of emergence if she intends her view to be plausible. However, the concept of psychological emergence cannot be physically respectable, even according to Stoljar’s object-based conception of the physical, which is supposedly liberal. Worse, with respect to the similarity between RM and physicalism, which is one of our main purposes of this section, the answer is just as obvious. With the addition of emergent qualitative properties and fundamental psychological laws to the world described by RM, it is no longer easy to see how such a world can be similar (theoretically and ontologically) to a world described by physicalism.

10.5. Naturalistic unacceptability

Even though Stoljar’s non-standard physicalism is of genuine philosophical interest, the primary interest of this dissertation is MN, not physicalism, and hence we should return to the former. Recall that physicalism is a version of MN, but not vice versa. Hence, physicalistic respectability and unacceptability implies naturalistic respectability and unacceptability, but not vice versa. We have seen that any plausible version of RM cannot be physically respectable and similar to physicalism, at least in the way Stoljar defines physicalism. But it does not imply the same with respect to MN. Nevertheless, I believe it can be easily shown that MN has the same relation with RM as that of Stoljar’s object-based physicalism (or ‘o-physicalism’, in his words).

As we have seen, combination without emergence is implausible. Thus, I shall focus on the relation between MN and combination by emergence. Recall I have developed two conceptions of MN in Part I, which are the minimalist thesis and the metaphysical paradigm thesis. Recall the minimalist thesis is as follows:

**The minimalist thesis:** Every entity or property instantiated in the actual world is metaphysically natural – (1) in the sense of (a) being posited by scientific concepts that are on the historical path of science or (b) being likely to appear on the path in the future, or (2)
in the sense of playing only a set of theoretical roles that is closely similar to one played by a metaphysically natural entity or property in the former sense.

The historical path of science is roughly the process of modification of scientific concepts over the course of history, along which they are modified and finally developed into the versions in modern natural science; and the doctrine is minimalist in the sense that it requires its follower to avoid positing anything that is not adequately related to the historical path of science, such as God, demons, and souls.

It seems to me that emergent qualitative properties and fundamental psychological laws are incompatible with this conception of MN, because they do not seem to be adequately related to the historical path of science. Two objections might be provided to my view. Firstly, many neuroscientists accept the approach called the neural correlates of consciousness (NCC), which (merely) attempts to discover the correlation between neural states and conscious states, and is neutral to what the actual constitutive relation between the two is like. With this in mind, emergent qualitative properties might seem to be an open possibility. Secondly, many scientists in different disciplines are emergentists who accept emergent laws. Hence, emergent properties like emergent qualitative properties and emergent laws like fundamental psychological laws may be adequately related to the historical path of science.

I will argue that some more careful considerations of the NCC and scientists’ openness towards emergentism will show that the objections are wrong. In response to the objection that the NCC and combination by emergence are consistent, it is important to note that the neutrality of the NCC to the nature of the actual constitutive relation between qualia and the brain does not imply that all possible constitutive relations are adequately related to the historical path of science. This is because the neutrality should be understood as a silence rather than an inclusion. For example, it is logically possible that the states of our souls (if there are any) have some stable correlations with our neural states. In this case, the NCC can be a successful neuroscientific method. Nevertheless, souls are not adequately related to the historical path. In sum, the NCC is merely silent on the issue of whether consciousness is realised by souls but is not positing souls. Similarly, the consistency between the NCC and combination by emergence alone does not show that emergent qualitative properties are adequately related to the historical path.

Hence, it is more crucial to assess the second objection and determine whether scientists’ openness towards emergentism will allow emergent qualitative properties and fundamental psychological laws to be adequately related to the historical path. I believe that the answer is no. On
the one hand, the fact that *some* (possibly) emergent properties, entities, or laws are referred to does not imply that *a particular* emergent property, entity, or law is referred to. It is true that even if some biological properties, entities, and laws are emergent ones, they are nevertheless directly referred to by the definition of MN. But it seems that the same simply does not happen to emergent qualitative properties and their laws. On the other hand, even if we allow (possibly) emergent properties or entities in science to be counted as a natural kind that is metaphysically natural, emergent qualia cannot be counted as a member of the kind. The reason is as follows. The (possibly) emergent properties in natural science are all results of *causal* emergence, in the sense that when there is a network of lower-level causal dispositions, the network possesses some causal dispositions that are inexplicable in terms of the lower-level causal dispositions. However, emergence of qualitative properties simply does not share this nature and hence cannot be counted as a member of the same natural kind.

We have seen that combination by emergence is inconsistent with the minimalist thesis. Let us move to the metaphysical paradigm thesis. Recall that it is as follows:

**The metaphysical paradigm thesis:** Every entity or property instantiated in the actual world is metaphysically natural – in the sense of being compatible with the metaphysical paradigm shared by the core theories (see Section 2).

Core theories are theories that have most if not all of their main concepts on the historical path; and these core theories share some metaphysical commitments, which I take to be a metaphysical paradigm. The metaphysical paradigm thesis is the view that MN is the paradigm, and its contents are the metaphysical commitments. In Section 3, I argued that these commitments include ‘there are no fundamental and nonlocal mental properties’, ‘there is a uniformity of everything’, and so forth.

It seems obvious why emergent qualitative properties and fundamental psychological laws cannot be metaphysically natural. The fundamental psychological laws are inconsistent with the commitment that ‘there are no fundamental and nonlocal mental properties’. Moreover, both emergent qualia and the fundamental psychological laws are inconsistent with the commitment that ‘there is a uniformity of everything’, because there is nothing we find in natural science that has similar nature to those things. As I have argued, combination by emergence has a nature that is significant from those of other kinds of emergence in natural science.
In sum, with the dilemma of physical combination, since combination without emergence is implausible because of the problem of appropriate mereological sum, the Russellian Monist must explain combination by emergence in order to have a plausible theory. But once there are emergent qualitative properties and fundamental psychological laws, the view can no longer be naturalistically respectable. Of course, we also want to determine whether the world described by the Russellian Monist can be similar to that of the metaphysical naturalist or not. But my answer is again no, since I simply cannot see how a world with emergent qualitative properties and fundamental psychological laws can be similar to a metaphysical natural world.

Returning to the implications of this section, I believe that I can safely conclude the following: (I) Regardless of whether the categorical acquaintance thesis is plausible or not, if it grounds on RM, then it cannot be said to be grounded on naturalistically respectable assumptions. This is because RM cannot be naturalistically respectable. (II) Any plausible version of RM cannot be similar to MN, and hence it is not obvious that it can share the theoretical advantages of MN over other versions of metaphysical non-naturalism.
Part III

Section 11. Qualia, the explanatory gap, and metaphysical naturalism

11.1. Introduction

In Sections 9 and 10, I argued that Russellian Monism (RM) is implausible and fails to be naturalistically respectable or similar to metaphysical naturalism (MN). Nevertheless, RM is only the (arguably) most attractive way to bridge the explanatory gap described by Joseph Levine (1983), according to which qualia have a non-causal nature that is intuitively known but not scientifically explainable. As we have seen in Section 8, what Bertrand Russell aims at is to propose an account that can bridge the gap, and what Daniel Stoljar aims at is to respond to the conceivability argument and the knowledge argument, which many believe to be, in turn, based on the gap (see, for example, Chalmers 2003a, p. 107).

As David Chalmers points out, our qualia appear to be not dispositional, causal, or structural (1996, p. 118). What this seems to imply is that they are categorical properties. If we were unable to bridge the gap, even if RM is wrong that such categorical properties are not those that take the role of other physical objects such as chairs and tables, it seems that they will always be a threat to the categorical ignorance thesis (CI) and MN. This is because qualia appear to be an idiosyncratic kind of categorical property that we can know of in our everyday experiences; and the historical path of natural science, which I argue is what characterises MN (see Section 2), does not seem to be able to reach them.

Providing an actual account of qualia is impossible in this dissertation because of the limit of scope, and hence I will have an alternative approach. What I will do is provide a very broad and general outlook of the scientific and philosophical study of qualia, in which several general strategies or ways to carry out research will be identified. Then I will determine which particular strategies might lead us into the gap or other dead ends, and which ones lead us out of it and back to the historical path of natural science. The idea of these strategies, which I call ‘meta-methods’, is as follows. By adopting a meta-method, a researcher explicitly or tacitly makes a decision about which feature or

---

56 This section is developed from a manuscript of an unpublished article I co-authored with Andrew James Latham, ‘Four meta-methods for the study of qualia’, in which I take the role of the first author.
role of her target of inquiry is the most important explanandum in her research project. In doing so, she also assumes a set of attitudes, background theoretical assumptions, research directions, methodologies, and so on, which correspond to the explanandum. More details will be provided as the discussion goes on. What I will argue by the consideration of meta-methods is as follows. Identifying a particular meta-method allows us to reasonably expect the explanatory gap can be bridged by some future scientific (or philosophical) studies that are on the historical path of science. This allows us to expect qualia to be metaphysically natural. Assuming that qualia are metaphysically natural, since probably no metaphysically natural things violate CI, we can also reasonably expect that qualia do not violate CI.

11.2. Qualia and the explanatory gap

Qualia, also called ‘phenomenal qualities’, cannot be easily defined. Different philosophers mean different things when they use the term. Daniel Dennett thus has the radical view that ‘philosophers actually don’t know what they are talking about when they talk about their qualia’ (2005, p. 88). There might be something correct in his view; but at this very beginning of the discussion, we can understand the word with more sympathy, and thus there is no need to assume that it is senseless. At the moment, I believe that it can be assumed that qualia are, as many philosophers suggest, ‘what it is like’ for a subject to have an experience. Setting the technical details and philosophical controversies aside, for most people, each conscious experience, such as that of seeing redness, that of tasting vegemite, that of being physically assaulted, consists of some ‘raw feelings’. As Jaegwon Kim points out, depending on whether these feelings are associated with pleasure or pain, they are what make our lives worthy of living, or terrible (2005, pp. 11-12). Hence, their existence is hard to deny.

For many philosophers, qualia resist scientific explanations. This is because they have a nature that is very different from those of all other entities or properties we know, resulting in the difficulty in explaining them by those other entities or properties. This is what the so-called hard problem of consciousness is about. One of the most influential discussions of the problem is provided in Joseph Levine’s seminal article ‘Materialism and qualia: the explanatory gap’ (1983), in which he proposes the so-called explanatory gap. In Levine’s view, natural science can only ever tell us about the causal roles of things. For example, the quale of ‘pain’, acting as a sign of bodily damage, may cause us to evade further physical harm. What natural science cannot explain is a ‘qualitative nature’ or the ‘phenomenal properties’ of qualia, that is, why the quale of pain ‘feels the way it does’ (p. 357, original
emphasis), since those properties non-causal. Levine does not intend to make a conclusion about which mind-body theory is true or false by the gap. What he attempts to formulate is an epistemological thesis instead of a metaphysical one (p. 354). More precisely, his view is that while it is possible that some physicalist mind-body theories are true, because of the gap we cannot know which particular theory is true (p. 359).

Nevertheless, not all writers endorsing this kind of argument are satisfied with this moderate conclusion. Many attempt to take the further step to reach the metaphysical conclusion that physicalism is false. Stoljar calls this more radical kind of argument the *structure and dynamics argument* (2006, p. 144; 2014, p. 29; 2015a, p. 325). It points out that physical properties are causal, dispositional, and/or structural, whereas qualia are not; from this, it argues that physical properties and qualia must be ontologically distinct. Its most influential contemporary proponent is Chalmers (1996, 2003a). However, a seminal version is already provided by the great 17th century German philosopher Gottfried Wilhelm Leibniz in his classical essay ‘The monadology’ (1714). The passage below is still frequently cited by contemporary philosophers of mind:

> We are moreover obliged to confess that perception and that which depends on it cannot be explained mechanically, that is to say by figures and motions. Suppose that there were a machine so constructed as to produce thought, feeling, and perception, we could imagine it increased in size while retaining the same proportions, so that one could enter as one might a mill. On going inside we should only see the parts impinging upon one another; we should not see anything which would explain a perception. The explanation of perception must therefore be sought in a simple substance, and not in a compound or in a machine. (p. 5)

In his thought experiment, Leibniz invites us to imagine a huge machine that can think and feel like a human brain. We can enter the machine; and yet, only mechanical motions can be seen inside it. Such mechanical motions cannot explain qualia. Analogically, the structure of our physical brain cannot explain qualia. Therefore, materialism is false; and dualism is true.

As Stoljar argues, one response we can provide to the structure and dynamics argument is that the *from-structure-only-structure thesis* it assumes (i.e. the thesis that structural, relational, or dispositional truths can only deduce further structural, relational, or dispositional truth) can hardly be shown to be true (2006, p. 151). In addition, it is noteworthy that, in my view, it is not obvious that science posits only structure and dynamics, since we have some possible candidates of non-
fundamental categorical properties irreducible to structure and dynamics, such as mass, size, and charge. This has been discussed in detail in Part II. Nevertheless, it is true that none of those candidates is anything like phenomenal properties, nor is it easy to see how posits of physics can constitute anything like phenomenal properties; and knowledge of neurophysiology does not seem to provide us any help. As Kelly Trogdon (2009) argues, even if we know little about H2O and general chemical laws, we would nonetheless have some grip on how nonwater physical truths can render water truth possible. But we have no clue on how this can be the case for qualia. This is what he calls the ‘really-not-having-a-clue feature’ of qualia (2009, p. 271).

However, as many writers correctly argue, this kind of argument is indecisive (Armstrong 1968; P. M. Churchland 1995; P. S. Churchland 2002; Stoljar 2006) because the really-not-having-a-clue feature might simply be caused by our ignorance. But then there is a dilemma facing the ignorance response. On the one hand, if we believe we are merely ignorant of neurophysiology as the Churchlands suggest, then, as Chalmers argues, the response cannot bear much weight, because there is an obvious disanalogy between the hard problem of consciousness and problems in other domains like the appearance of the existence of vital forces. Precisely, it is more or less obvious that the matters in those domains are usually functions and require functional explanations, whereas qualia are not (1997, p. 381).

On the other hand, if we have a liberal conception of what we might be ignorant of, the challenges the hard problem provides to us may, in fact, persist. Consider Stoljar’s Nagelian non-standard physicalism, which we have seen in Section 8. In this view, a currently unknown type of experience-relevant nonexperiential truth that ideal physics will posit can account for qualia. A worry about the view is, I believe, that it is overly liberal and too easy to be true if no exact detail is provided with respect to what the type of truth is like. Recall that we have discussed Hempel’s dilemma in Section 2.4.1, in which one horn is that if we define the physical with reference to ideal physics, then it is trivially true. This is because ideal physics might be understood as a set of theories that has successfully described everything (in the actual world), and thus everything (in the actual world) could be accordingly counted as physical. Of course, Stoljar makes it clear that the Nagelian non-standard physical truth he has in mind is nonexperiential, which is, indeed, an important constraint. But it is still too easy for something to be counted as physical, and hence it is too easy for Stoljar’s Nagelian non-standard physicalism to be true, even though it should not be said to be trivially true.

In fact, such radical liberalisation of the concept of the physical seems to be what Stoljar purposefully brings about, because the examples of the Nagelian non-standard physical he discusses
in length are all extraordinary things that we might be ignorant of because of some metaphysical limitations our current science has not overcome, not merely because our current science is incomplete. Examples include RM and the evolutionary limitation of our cognitive abilities to know about things. On the contrary, even though he briefly mentions that ignorance of neurophysiology is a possibility (2006, p. 71), no consideration is actually given. This implies that, contrary to the theoretical modesty of Nagelian non-standard physicalism, the Nagelian non-standard physical truth that he seems to *actually and implicitly* expect is something that has a considerable distance from the historical path of natural science, not anything similar to the posits of current neuroscience. This threatens MN and CI. The reasons are as follows respectively. For MN, because the doctrine, as I have argued in Section 2, should be characterised by the historical path of science, the existence of something that has a considerable distance from the path is likely to be inconsistent with the doctrine. For CI, what Stoljar has in mind makes it difficult to predict whether the categorical acquaintance thesis – according to which we can be acquainted with categorical properties through acquaintance with qualia – is true, which is inconsistent with CI. This is because without the historical path of science, which we can use as a reference, we would have difficulties in predicting what kind of Nagelian non-standard physical fact – consistent with the categorical acquaintance thesis or not – will turn out to be the correct explanation of qualia.

One option to overcome these worries is, of course, to accept eliminativism about qualia. In fact, eliminativism is considered by Levine as only one single way that the physicalist can bridge the gap. But, as he correctly points out, ‘this kind of intuition about our qualitative experience seems surprisingly resistant to philosophical attempts to eliminate it’ (1983, p. 361). What Levine presents is a dilemma to us. According to the dilemma, we can either respect our phenomenal life or stay on the historical path of science, but we can never have both.

My own attitude towards the nature of mind can be described as ‘scientific’. In line with this attitude, my approach to bridging the explanatory gap and solving its challenges to MN and CI is as follows. It is fairly obvious that neuroscience has not discovered every fact about qualia. Hence, to account for qualia without violating MN and disrespecting our phenomenal life, it seems to me that the best option is to accept what Stoljar calls Nagelian non-standard physicalism about qualia – which appeals to an expectation that future science could explain qualia – but not to accept eliminativism. Nevertheless, in order to avoid violating MN, having Nagelian non-standard physicalism is not sufficient. There is also a need for the expectation that the Nagelian non-standard physical fact is not, as Stoljar seems to implicitly expect, anything distinct from the historical path of science but is, as
the Churchlands advocate, something like a fairly ordinary neurophysiological fact. Put simply, what we need is an expectation that neuroscience (or some related sciences like psychology, cognitive science, and so on) can bridge the explanatory gap without diverging from the historical path and disrespecting our phenomenal life. If the expectation can be reasonably made, the potential threat of qualia to CI will dissolve as well. With the assumption that qualia are metaphysically natural and could be explained by something like ordinary neurophysiological facts, it can be expected that they probably cannot allow us to be acquainted with categorical properties and hence probably will not violate CI. The matter of concern is then how the expectation can be reasonably made when qualia, as Chalmers points out, (are believed to) have some features that resist neuroscientific explanations. In the rest of this section, I will argue that the scientific and philosophical study of qualia can, and ought to, be carried out in a way that can overcome the difficulties facing the expectation. If my arguments are successful, then the explanatory gap argument is not compelling. In addition, we can justifiably have the expectation that the correct understanding of qualia could be consistent with MN and CI.

11.3. Meta-methods for the study of qualia

What I am going to argue is as follows. Research projects of qualia can be categorised as adopting different ‘meta-methods’. By adopting a meta-method, a researcher – intentionally or unintentionally, and explicitly or tacitly – makes a decision about which feature or role of her target of inquiry is the most important explanandum in her research project. In doing so, she also assumes a set of attitudes, background theoretical assumptions, research directions, methodologies, and so on, which correspond to the explanandum. The choice of meta-method, I will argue, may significantly influence her findings or conclusions and, in particular, whether they could bridge the explanatory gap, respect our phenomenal life, and stay on the historical path of science.

Differences in meta-methods can occur in a range of research areas, not only that of qualia. Let me illustrate by example. Consider studies of God by religious believers. In the areas of philosophy of religion and theology, there will always be some difference in findings contingent on whether the religious researcher – who might be a theist philosopher of religion, a theologian, or whatever – studies God as ‘the God of the philosopher’ or ‘the God of the believer’. In the former case, God is assumed to be a posit of classical metaphysics and takes the roles of the first cause, a necessary being that everything else grounds on, and so on. In the latter case, God is supposed to be a personal being, which the believer can enter into a personal relationship with. Unsurprisingly, because of these
differences, even though these researchers might assume that they are directing their inquiries onto the same target, their research projects will consist of very different sets of attitudes, background theoretical assumptions, research directions, methodologies, and so forth. I adopt the term ‘meta-method’ to describe this phenomenon because it is more general, broader, and less precise than any formal research method (e.g. methods of measurements, analysing data, etc.).

There will always be a lot of overlap between different meta-methods, and the dividing lines between them may always be vague. After all, a comprehensive account of a subject has to consider all its possible features and roles, not just the features or roles that particularly interest us. Even when one feature or role is more relevant to us than another, it is not always easy to classify things strictly as the features or roles of the target. Consider the study of the God of the philosopher. If some philosopher of religion determines that God is a conscious, omnibenevolent, creator entity because of some metaphysical considerations, it would seem to follow that developing a personal relation with people is the sort of act a conscious, omnibenevolent creator entity would do. Thus, in some sense, the God of the philosopher might have many things in common with the God of the believer. With this in mind, what meta-method a researcher actually adopts is often a relative issue, not anything absolute. Meta-methods are not inconsistent with each other in the sense that if two researchers adopt different meta-methods, then they are necessarily in disagreement with each other. A researcher may shift back and forth from one meta-method to another when she works on different research projects and may even adopt different meta-methods when she is working on different components of the same research project. But she remains wholly consistent. However, differences in meta-methods may still lead researchers to develop very different findings and conclusions. As in the case of God, scholars who believe in the same religion and (supposedly) the same God can have tremendously different conceptions of her because of such differences.

In the rest of this section, I will survey the different meta-methods for the study of qualia and will argue that differences in meta-methods can result in substantive differences in the findings of those adopting them. I will identify and describe four meta-methods for the study of qualia. Specifically, they are the theory-centred meta-method, the property-centred meta-method, the argument-centred meta-method, and the event-centred meta-method. I will illustrate each meta-method by providing examples of scientists and philosophers investigating qualia who we think exemplify these meta-methods. Then, I will highlight the influences of each meta-method and argue that those other than the event-centred meta-method are not as well suited to developing a scientific conceptualisation of qualia that is on the historical path. I will argue that the theory-centred and the
property-centred meta-methods are inflexible and can lead us to a false dilemma between extremist affirmative and eliminative attitudes towards qualia. I consider it a false dilemma because some intermediate – and more scientifically useful and constructive – positions are available if we adopt the event-centred meta-method. To defend such positions, I will sympathetically address the views about qualia of some eliminativists influenced by a Feyerabendian or similar conception of the history of science, such as Patricia Churchland (1988) and neuroscientist Stanislas Dehaene (2014). While they are right to believe that the history of science encourages us to discover progressive theories, not conservatively retain old dogmas, this does not require us to abandon everything we currently have.

Finally, I will argue that the event-centred meta-method is the most promising because it allows us to revise our ontological commitments and have a flexible methodology. Put simply, the idea is that our initial ontological commitments regarding the nature of qualia should be a base for revisions to build on, as opposed to dogmas to hold or abandon. In addition, we should also not hold any methodological absolutism concerning how we should inquire into the nature of qualia. Instead, we should allow each possible phenomenological method to prove its worth by providing empirical or philosophical findings, not simply by some very abstract a priori reasoning that is detached from our findings. I will argue that the developmental strategies allow the study of qualia to be developed upon the historical path of science in an easier way.

11.4. The four meta-methods

In this subsection, I describe four meta-methods and provide examples of researchers who I consider to be motivated by those meta-methods when developing theories about qualia. To narrow down the scope of investigation, like many contemporary scientific researchers, I assume that physicalism is true and focus only on research programs that assume the same. Of course, some of our proposed meta-methods are not exclusive to physicalist research programs. However, for the purposes of this dissertation, I shall set aside any further discussion of this matter.

There is a need to re-emphasise that the meta-methods are general and broad. A meta-method is the set of theoretical and methodological assumptions that are made because of a researcher’s decision about what feature or role of the subject is her target of inquiry. Hence, they have lots of overlap and no clear dividing lines between them. The four meta-methods we describe closely relate to one another, and the examples of researchers or research projects we offer up are contestable. However, the main purpose of this section is not to provide a strict set of categorisations that everyone can fit
neatly into, but rather to identify some major research streams of the study of qualia and reflect on them in the later parts of this section.

11.4.1 The theory-centred meta-method

Researchers who adopt a theory-centred meta-method are particularly interested in how qualia play the role of a theoretical entity picked out by our folk psychological theories. Accordingly, these researchers then attempt to translate and/or reduce these folk theories into theses that are compatible with their academic traditions (such as neuroscientific theories or functionalist explanations).

One seminal example of this meta-method is David Lewis’s (1995) analytic functionalism. In this view, ‘qualia’ is the name given to the occupants of a particular causal-functional role described in our tacitly held folk psychology. Qualia simply are, then, whatever plays that causal-functional role. If physicalism turns out to be true, then there may be no perfect occupants of the role, but there may still be some imperfect ‘next-best deservers’ (p. 140). Lewis identifies part of a quale’s causal-functional role via what he calls the identification thesis. According to the thesis, when one has a quale, one knows all the essences of it (p. 142). Lewis believes that the full version of this thesis is unacceptable to the physicalist, but he believes that the physicalist could accept it in ‘some not-so-demanding everyday sense’ (p. 144) – for example, one could know the relations of acquaintance one bears to qualia (for further details, see Lewis 2004 and the so-called ‘ability hypothesis’).

Eliminativists such as Paul Churchland (1985), Patricia Churchland (1988), and Alex Rosenberg (2011) also appear to be examples of researchers who have adopted a theory-centred meta-method. They argue that since our folk psychological theories are shown to be radically inaccurate by science, any theoretical entities posited by such theories, like qualia, probably do not exist. While intuitive conceptions of qualia may one day reduce to some lower-level account (e.g. neurophysiological explanation), there is no guarantee this will occur. This is because being reducible requires a theory to be ‘basically’ correct, and since our folk conceptions – including the theories or the so-called immediate introspections – are likely to be radically flawed, they should simply be replaced by future scientific knowledge (1988, p. 301).

Some researchers are not eliminativists about consciousness but, nonetheless, hold an eliminative attitude towards qualia. For neuroscientist Stanislas Dehaene (2014), the concept of qualia is a hypothetical one based on our ill-defined intuitions. What qualia refer to is ‘pure mental
experience detached from any information processing role’ (p. 262). However, once our intuitions have been improved through advances in neuroscience and computer simulations, the concept will simply vanish and be considered prescientific (e.g. a fate similar to other prescientific concepts such as vitalism).

11.4.2 The property-centred meta-method

Researchers who adopt a property-centred meta-method understand qualia to be bundles, or bearers, of some metaphysical properties observed through introspection (often, by those same researchers). They then focus on how to situate them within the natural world (if possible). This meta-method is very similar to the theory-centred meta-method in the sense that both are based on our intuitions about qualia. However, the property-centred meta-method does not assume any comprehensive folk psychological theory; instead, it takes those metaphysical properties to be knowable by direct, immediate and pretheoretical observation of our experiences through introspection. An analogy can be drawn with the scenario of seeing a turtle in front of oneself along with all its features. Many consider the knowledge about them to be too obvious and immediate that it is unguided by any theoretical reflection. As Galen Strawson (2008b) puts it, ‘we have it and know it at every waking moment’ (p. 53), and ‘nothing in this life is more certain’ (p. 21). Different researchers disagree about which set of properties is the right one because of differences in their introspection. Commonly mentioned ones include being non-causal, intrinsic, ineffable, incorrigible, and so on.

Levine’s work on the explanatory gap (1983) appears to reflect the property-centred meta-method. As we saw earlier, the crucial point of his view is that natural science cannot explain a ‘qualitative nature’ or the ‘phenomenal properties’ of qualia – that is, why the quale of pain ‘feels the way it does’ (p. 357, original emphasis) – since those properties non-causal, as opposed to physical properties, which he takes to be essentially causal. This seems to be heavily based on Levine’s own introspection. As he states, what he describes is an important intuition he finds hard to deny (p. 361).

Russell,\(^{57}\) holding that qualia are ‘intrinsic’ and non-causal, proposed RM. As we have seen in Section 8, in his view, physics can only tell us about the causal and spatiotemporal properties of physical entities, not their intrinsic properties, whereas we can only know of the intrinsic properties

---

\(^{57}\) It is noteworthy that Russell endorses different views at different stages in his life. See Tully (2003) and Wishon (2015) for detailed studies of his different views on qualia.
of the world in our perception. Thus, the mind-body problem can be solved by considering qualia to be the intrinsic properties of the physical (1927b, p. 154). It is important to note, though, that Russell was a neutral monist and explicitly denied being a materialist (1992/1927a, p. 382). Despite this, RM has attracted many physicalist followers who have attempted to develop physicalist versions of this view, including Stoljar (2001), Galen Strawson (2008b), and Barbara Montero (2015).

In his influential article ‘Quining qualia’ (1993a), Dennett proposes an eliminativist view of qualia that exemplifies the property-centred meta-method. Distinct from the Churchlands (who adopt the theory-centred meta-method), Dennett begins with the assumption of property-centred meta-method and considers qualia to be a set of ‘special properties’ (p. 382) that, ‘in one form or another, is “obvious” to most people – to scientists, philosophers, lay people’ (p. 382). The properties he spells out include their being ‘ineffable, intrinsic, private, directly or immediately apprehensible in consciousness’ (p. 385). Then, he argues that these properties are, in fact, inconsistent with our intuitive assumptions about them. Qualia do not exist because those who talk about qualia do not know ‘what on earth they are talking about’ (p. 382).

11.4.3. The argument-centred meta-method

Researchers who adopt the argument-centred meta-method are usually philosophers defending physicalism. They are interested in the role qualia play in the arguments against physicalism that appeal to our first-person experiences (e.g. Jackson 1982; Chalmers 1996). Usually with the intention to defend physicalism, these researchers offer some conceptual analyses that are intended to identify features of qualia that, in their views, lead to the confusion of the non-physicalist.

Of course, the arguments against physicalism these researchers are responding to might also exemplify meta-methods that could be classified as either theory-centred or property-centred. Consider Frank Jackson’s\(^{58}\) (1982) influential Mary thought experiment. In this thought experiment, we are invited to imagine a scientist, Mary, who has access to all the physical facts relating to the experience of colour but has spent her whole life in a room, only ever experiencing black and white. Once Mary leaves the room and sees something red for the first time, it appears that she has learned something new. However, because Mary had access to all the physical facts about colour experience, this new fact cannot be physical. This led Jackson to conclude that qualia are not physical (p. 130). Jackson’s strategy relies on the intuition that Mary learns something new about the experience of

\(^{58}\) Frank Jackson no longer holds this view.
colour. While he does not explicitly tell us the source of this intuition, we might consider the intuition as arising from some tacitly assumed folk psychological theory of qualia or tacit awareness of the metaphysical properties possessed by qualia. Respectively, these lines of thought are congruent with the theory-centred or the property-centred meta-methods.

Researchers who adopt the argument-centred meta-method, however, are not those who are sympathetic to the arguments against physicalism but those who reject them. More specifically, these researchers are merely interested in replying to arguments against physicalism by using conceptual analysis to uncover some features of qualia that, in their views, lead to the confusion of the non-physicalist philosopher.

John Bigelow and Robert Pargetter (1990, 2006) are an example of researchers who have adopted an argument-centred meta-method. Their acquaintance hypothesis is a reply to Jackson’s Mary thought experiment and aims to explain how Mary can acquire new knowledge without committing to non-physicalism. According to their hypothesis, we can hold different acquaintance relations to the same fact by causally relating to it in different ways. Mary obtains new knowledge not by learning a new fact but, instead, by developing a new acquaintance relation to a fact she already possesses. They conclude that ‘although qualia are physical, people cannot know all there is to know about them unless they experience them for themselves’ (1990, p. 147). For them, Jackson is correct that Mary acquires new knowledge, but his understanding of what the new knowledge consists of is a mistake.

11.4.3. The event-centred meta-method

Researchers who adopt the event-centred meta-method understand qualia as phenomena that must be explained by scientific and philosophical theories. The term ‘phenomena’ here is meant in a scientific sense, designating events that are susceptible to scientific exploration. What differentiates this meta-method from the theory-centred meta-method and the property-centred meta-method is the notion that the main explananda are the neurological events where qualia are taken to exist, not our folk psychological theory of qualia or some set of metaphysical properties we attribute to qualia through our introspection. In this sense, much like other natural events such as typhoons or biological reproduction, what interests the researcher is not our intuitive or initial concept(s) of the events of concern, but their hidden natures, which we have to uncover empirically.

No doubt, there is an inevitable commitment to either the folk psychological theory or some list of introspected properties at the beginning of the investigation. The event-centred meta-method
requires a starting point at which qualia are provided an initial description, thereby allowing researchers to begin their inquiry. Without such a starting point, the inquiry has nowhere to begin. However, the researcher sees the folk psychological theory or the list of introspected properties as what we initially attribute to the explananda, or in other words, initial data and hypotheses about the explananda, not the explananda themselves. Making use of these data and hypotheses to begin the inquiry does not mean the researcher has to consider them to be true. The correct way to describe these events is an empirical matter. Eventual revisions in the form of more advanced theories and refined lists of properties are always possible. While these researchers may make methodological use of other meta-methods, they are not theoretically committed to them. After all, this is much like the case of physics. We must first have some intuitive or initial concepts of matter for inquiries into it to be possible, but then we uncover the events behind our concepts in the empirical world and revise those concepts in this process.

Philosopher Jakob Hohwy and psychologist Chris Frith (2004) are examples of researchers who we associate with the meta-method. In response to the question ‘Can neuroscience explain consciousness?’, they describe a particular kind of consciousness that they believe neuroscience does provide a successful explanation for. By drawing on a wide body of evidence, including introspective reports, neuroscience, and psychiatric studies, they argue that our feeling of being in control can be explained by our predictions for our own actions. No conceptual frameworks of qualia, such as introspective reports or their interpretations, are taken to be true until they show their explanatory superiority over others (p. 182). They do not stop at merely drawing correlations between qualia and neural states (i.e. the neural correlates of consciousness [NCC]). That is, they not only aim to tell us when and only when we have a sense of control, but also what that sense is. They acknowledge that, at present, their method cannot bridge the explanatory gap completely but argue that the problem is more likely to be solved through many small research steps rather than any single discovery (p. 196). We take this to be an example of the event-centred meta-method since no initial conceptions of qualia are taken to be true. Instead, qualia are considered events whose natures must be discovered bit by bit.

Some researchers appear to have tacitly held attitudes consistent with the event-centred meta-method. Neuroscientists Gerald Edelman and Giulio Tononi (2001) classify qualia as a natural phenomenon and subject worthy of scientific investigation. They warn us about the limitations of philosophical efforts to reveal the nature of qualia. The nature of qualia, much like those of other natural phenomena, is not the kind of thing that is uncovered from the armchair; it requires scientific
observations and experiments (p. 6). They also note that ‘no scientific description or explanation can substitute for the real thing’ (p. 12), and that ‘being comes first, describing second’ (p. 15). Put simply, Edelman and Tononi remind us that qualia as events are distinct from theories of qualia. What we can do, such as when we empirically investigate other natural phenomena, is propose scientific explanations drawn from a wide range of methodologies that explain both the phenomenon’s properties and the conditions under which it obtains.

Edelman and Tononi’s research program begins by spelling out the properties of qualia they wish to explain by observing our most general experience: specifically, its privacy, unity, and endless variety (p. 20). This starting point is similar to the property-centred meta-method. However, their choice of properties is preceded by their scientific assumption that qualia are physical and evolved (pp. 12-16). Our initial understanding is not to be taken for granted and must accommodate other scientific findings in order to be assumed correct. For example, as they develop their theory of qualia, they reject atomism, according to which each quale is an element independent of each other (pp. 162-163) – even though atomism is attractive from the perspective of introspection (p. 162). I believe that Edelman and Tononi’s view of qualia as events means that they do not confuse its presumptive set of properties as its ultimate explanandum and, consequently, do not see presumptions as non-negotiable. Rather, they see the set of properties as a useful collection of hypothetical posits that must be carefully selected. It is from this tacit method that they developed their scientific theory of qualia: the dynamic core hypothesis, which has attracted many followers (for more details, see Edelman & Tononi 2001).

11.4.5. Heterophenomenology

Some readers may wonder why I have not included Dennett’s (1991, 1993b, 2003) influential heterophenomenology as a distinct meta-method. Heterophenomenology is a method of phenomenology. In this method, the introspective data provided by participants are considered no more than speech acts or texts. These texts are taken as raw experiential data whose contents are not assumed true but fictional (1991, p. 78). As Dennett writes, ‘The reports are the data – they are not reports of data’ (1993b, p. 51). What the researcher then does, from a third-person perspective, is search for causal explanations of the participant’s reports; she never enters a first-person perspective or makes any claim from her own introspections (2003, p. 21).

Dennett argues that this method applies to reports on qualia (1991, p. 407). To restate it as a meta-method for the study of qualia, qualia are assumed to be nothing more than a kind of content in reports provided by experimental subjects. They are not considered to be real, and the researcher is
only interested in discovering what causes participants to (mistakenly) make such claims. I do not wish to include heterophenomenology as a distinct meta-method, the reasons for which will become clear as this section progresses.

11.4.6. Interrelations between meta-methods

I have noted that meta-methods are closely interrelated and have lots of overlap between them. This is because: (1) considering one feature or role of qualia to be the most important explanandum does not prevent one from investigating others, (2) one feature or role might be closely related to another, and (3) the dividing line between different features or roles are vague.

For example, those who have adopted the theory-centred (not theory-only) meta-method do not need to have an exclusive concern for folk psychology. It is possible to think that many of our folk psychological theories develop from our immediate introspection of the properties of qualia. Those who have adopted the property-centred meta-method do not assume a theory as sophisticated, detailed, and comprehensive as the folk psychological theory used by the theory-centred meta-method. Still, those claims of introspected properties could be argued to resemble ‘small theories’ held by the folk. In addition, if the immediateness and theoretical reflectiveness of introspection is a matter of degree, then the dividing line between an immediateintrospection and a folk psychological theory might be vague. The arguments against physicalism considered by those with an argument-centred meta-method, as we have already noted, are reliant on folk theories and intuitions about qualia’s properties. Similarly, the event-centred meta-method, as we already have noted, is reliant on folk theories and introspected properties of qualia in order to point out instances of qualia to begin investigating (even though they do not have to be assumed to be true). In this sense, the event-centred meta-method can be considered as an extension of the theory-centred and the property-centred meta-methods.

One can also shift from one meta-method to another across time. For example, Patricia Churchland (1986, 2002) sometimes appears to be adopting the event-centred meta-method rather than the theory-centred meta-method. Aside from suggesting that qualia do not exist, she sometimes appears to be attempting to provide scientific explanations of qualia that replace the folk’s.

11.5. An assessment of the theory-centred and the property-centred meta-methods

In this subsection, the theory-centred and the property-centred meta-methods are assessed together, as we argue that they share the same major weakness. Other meta-methods are assessed separately in subsequent sections. I argue that both the theory-centred and the property-centred meta-methods are
disadvantageous to the development of the study of qualia – particularly its following of the historical path of science. Since René Descartes, there are countless researchers who consider (or appear to consider) their current understanding of qualia to be (almost) incorrigible (e.g. Searle 1990; Chalmers 2003b; Strawson 2008b). However, it should now be common knowledge that many of our initial assumptions about how consciousness operates are mistaken. Accordingly, it is difficult to see why those about qualia should be exceptions. It might be true that our everyday experience appears to convey to us that qualia exist, but it remains difficult to see why our contemporary or intuitive views about qualia should be accurate. Placing our assumptions at the centre of the inquiry prevents flexibility and revisions to research directions if any assumption turns out to be wrong.

In the cases of research programs that adopt the theory-centred meta-method, if folk psychology turns out to be radically false, it is difficult to see how they could go on. The Lewisian methodology does, indeed, consider folk psychology to be false because the qualia role picked out by the theory, according to Lewis, has no perfect occupant. However, if folk psychology turned out to be radically false, then there may be nothing that even imperfectly fits the qualia role. I do not disagree with Lewis’s approach to the identification thesis, but the qualia role could still be radically false. This is because the identification thesis might not be the only thesis our folk psychology assigns to the qualia role. For example, Dehaene suggests that qualia are believed to be ‘pure mental experience detached from any information processing role’ (2014, p. 262). These other theses might be radically false. Moreover, Lewis might also be assuming too much when he seems to consider folk psychology to be a consistent theory. It may actually be, as Dennett suggests, inconsistent in some important ways. Furthermore, as Patricia Churchland (1988) argues, if a kind is not a natural kind, it should be eliminated in our scientific ontology; and, indeed, our folk psychology about qualia might commit to non-natural kinds. Her argument will be discussed in detail in the coming paragraphs.

---

59 For an overview of scientific findings that show the inaccuracy of our initial assumptions about consciousness, see P. S. Churchland 1988, p. 288-290; Dennett 1991, p. 68-70. I shall not repeat the cases here.

60 Certainly, what is meant by a causal role is a bit unclear, and thus what it takes for a causal role to be true (or false) is also a bit unclear. Therefore, it could be asked under what conditions the qualia role could be radically false. For example, if a causal role is defined only as whatever it is that causes our beliefs about the role, this would almost guarantee the existence of qualia to be trivially true, since the only way that it could turn out to be false is if our thoughts about qualia are causeless or causally inefficacious. But qualia are not supposed to be so; zombies, if possible, have thoughts about qualia which are not causeless and causally inefficacious, but they do not have qualia. A clue might be Lewis’s causal descriptivism, which he proposes elsewhere (1984). In the view, the references of names have to be determined by descriptive fitness, while descriptions have to be couched in causal terms. With this in mind, a causal role seems to be a comprehensive description of the properties of its reference in causal terms. In this case, the qualia role could be radically false if some important causal descriptions of it do not correspond to its actual features.
On the other hand, philosophers and scientists who have adopted the property-centred meta-
method do not assume a theory as sophisticated as folk psychology, yet the assumption that qualia
must possess the particular set of introspected properties seems to, in a similar way, lead the study of
qualia into a dead end and, often and worse, to diverge from the historical path of science. More
precisely, authors with such persistence are often led to unnecessarily radical conclusions and, in
particular, mysticism about qualia that resists scientific inquires in order to make sense of the
introspected properties. For example, Russell is driven to believe that qualia are some kind of intrinsic
characters of physical entities that are beyond the reach of physical science. Similarly, Levine, who
is a physicalist, is driven to believe in the explanatory gap, according to which it cannot be known
which physical explanation of qualia is true.

We have seen that the inflexibility of the theory-centred and the property-centred meta-methods
lead to either a (probable) failure or mysticism about qualia; and both are de facto dead ends for the
development of a scientific conception of qualia that follows the historical path of science. Hence, it
is understandable that many authors who understand qualia via the lens of the two meta-methods are
drawn to eliminativism about qualia. In fact, when Levine writes his seminal article on the
explanatory gap, what he has left us with is the dilemma that we have to accept either the explanatory
gap – which is a version of mysticism about qualia – or eliminativism – which he does not consider
to be tenable (1983, p. 361). Nevertheless, before any judgment concerning the dilemma is made, we
should assess the nature of eliminativism carefully and sympathetically. What I am going to argue is
as follows. Eliminativists such as the Churchlands, Rosenberg, Dehaene, and Dennett may be too
quick to deny the existence of qualia due to a lack of flexibility inherent in the theory-centred meta-
method they share with Lewis. From the history of science, we can learn that even if an old theory is
radically false, this fact should not automatically imply that things of concern in the theory do not
exist. This also provides us a way out of the dilemma Levine offers us.

A major argument for eliminativism is offered by Patricia Churchland who adopts the theory-
centred meta-method. In her argument, she makes use of resources from Paul Feyerabend’s
incommensurability thesis in philosophy of science, which is (roughly) the view that translations
between old and new theoretical frameworks are improbable if not impossible (P. S. Churchland 1988,
p. 275; see also Feyerabend 1981). On her view, it is possible for there to be a reductive explanation
of mental phenomena (including qualia) through neurobiology, in a similar sense to the case that light
is reduced to electromagnetic radiation. In this case, there will be an identification of mental
phenomena and neurobiological phenomena, and science can preserve the concepts of mental
phenomena. However, for this kind of case to happen, the old theories have to be, in her words, ‘pretty much correct’ (301). Such cases, she argues, are atypical in the history of science; and in most cases smooth reductions is impossible because, as the incommensurability thesis tells us, old and new theories are incommensurable. As psychology and neurobiology co-evolve, she concludes, mental phenomena should probably be given up like caloric fluid and vital force.

I am sympathetic to this argument because, despite some philosophers’ contention that eliminativism is simply an implausible denial of obvious phenomenal facts (e.g. Chalmers 1996; Strawson 2008b), what Churchland, in fact, argues is that she expects some more advanced accounts of those facts, even though those accounts probably would not make use of the concept of the phenomenal. After all, it is difficult to see why replacements of old theories through scientific progress should not be expected, unless one holds the strong philosophical notion that our current folk psychology is incorrigible and complete. Nevertheless, I believe that there is a way to understand what the history of science teaches other than Churchland’s. More precisely, there are many more ways to preserve the elements of old theories than she considers, even in cases where new theories have replaced old ones. No doubt, she is correct that eliminations like that of vital force are always possible, but we cannot know a priori whether they will actually happen. Hence, she might be too quick to dismiss the elements of our folk psychology.

Let me illustrate. Churchland provides two scenarios in which the incorrectness of the old theories might consist. One is when the theoretical contents are simply false (1988, pp. 287-290) and the other is when the described kinds are not natural kinds but nominal kinds (pp. 284-287). However, in the history of science, cases of both kinds have often undergone revision without elimination. For example, theoretical entities such as gravity and atoms have been given new meanings as physical theories develop. <Gravity> in Newtonian physics was revised before it developed into its counterparts in string theory or the theory of relativity. Atoms are no longer considered indivisible or like billiard balls, as they have been considered historically. <Planet> and <acid> in older theories have been proven to be non-natural kinds, but they were then revised and given newer meanings and newer sets of inclusions in newer theories. This kind of revision provides a basis for new theories to develop upon. It also makes the development of scientific theories easier because the new theory can make use of the rich theoretical resources already present in the old theory.

Scientific practice is pragmatic and realistic; constructing every element of the new theory by itself is not typical. In fact, not only have the theoretical contents and natural kinds found in scientific theories undergone revisions, so too have intuitive and ancient concepts about ordinary physical
entities such as water, earth, and fire – which, by the light of current physics, have all been shown to be radically false. The early 20th century physicist Arthur Eddington famously and correctly notes that physical objects according to the folk theory have extension, permanence, colour, and substantiality. He then reminds us that according to current physics, they are actually empty spaces with fields of forces (1928, pp. 5-9). Nevertheless, scientists and philosophers (with the exception of mereological nihilists) rarely consider them to be in need of elimination and frequently use them in scientific and philosophical theories. Given all these other cases, it is difficult to see why <qualia> should be an exception whereby the possibility of revision remains open. Churchland and Dehaene are correct that an elimination of qualia like that of vital force is always possible, but we cannot know a priori whether it will actually happen, and hence the analogy is not conclusive.

This, of course, involves a debate in philosophy of science on the connection between theoretical contents or natural kinds and their revised counterparts. The issue remains controversial. Some consider them to share the same reference through causal theories of reference (e.g. Kitcher 1993) or structural similarities (e.g. Schurz 2009); others consider the revised counterparts to be merely borrowed concepts (e.g. Kuhn 1996). However, I take no stance on this controversy. My consideration is merely pragmatic, based on how scientific progress actually happened in history. We have seen how scientists (and philosophers and the folk) revised other concepts with the aim of allowing them to correctly correspond to the event they are tracking, and it is simply difficult to see why the same could not happen for <qualia>. Some readers might consider my position to be eliminativism in disguise, because the old concept of qualia, after having undergone revisions I expect, is nevertheless abandoned. I disagree. It is important to see that revisions have a gradual evolutionary nature and are thereby not simple abandonments as Churchland has in mind.

Eliminativists who apply the property-centred meta-method fare no differently from those with the theory-centred meta-method. We have seen that Dennett argues that the ‘obvious’ properties of qualia are inconsistent with our intuitive assumptions about them. As a result, he simply concludes that the concept of qualia is a confusion, and therefore qualia do not exist. The quick denial of qualia’s

---

61 No doubt, some might worry that all radically false old concepts do not track anything. It is impossible to address this worry in detail in this section because it involves debates in the theory of reference that are far beyond our scope. But empirically speaking, it seems to be fairly obvious that old theories such as Aristotelian and Newtonian physics are tracking some phenomena, despite the fact that they do not correctly correspond. In particular, to assert that old concepts do not track anything requires us to believe that our ancestors’ worldviews track almost nothing, which is quite implausible. Hence, it seems that we have good reason to assume that old concepts do track something, not the radical opposite view. In fact, even in the recent literature on eliminativism, more eliminativists endorse scientific eliminativism, according to which folk concepts should not have any place in our scientific ontology; and semantic eliminativism, which is the more radical view that folk concepts do not refer to anything, is less popular.
existence, like Churchland’s, is a result of holding inflexible assumptions and neglecting the possibility of revision. Recall that Dennett’s heterophenomenology assumes that there is no phenomenology but only propositional attitudes. This is a result of his eliminativism, which is motivated by a property-centred meta-method. As Dennett explains, ‘There is no such phenomenon as really seeming – over and above the phenomenon of judging in one way or another that something’ (1991, p. 364), and ‘[speech acts are] sufficient to produce heterophenomenology’ (p. 365).

We have seen why the dilemma between extremist affirmative and eliminative attitudes towards qualia is a false one led by the inflexibility of the theory-centred and the property-centred meta-methods. More precisely, an intermediate position is possible via revision of our concepts. I believe that neglecting this fact is a disadvantage because it blocks the kind of revisions typical in other scientific fields, which is a useful way for researchers to develop new hypotheses that can bridge the explanatory gap without disrespecting our phenomenal life and diverging from the historical path of science. For example, if we take Dehaene’s (2014) advice to wait for the concept of qualia to vanish, then all we can do in the interim is focus our studies on other aspects of consciousness; and if we accept Dennett’s heterophenomenology and thereby assume that there is no phenomenology but only propositional attitudes, it seems that we will lose a lot of valuable materials and insights that might otherwise be seminal in suggesting future studies of consciousness that could be on the historical path of science. Worse, both do not seem to be able to respect our phenomenal life; and this makes them unattractive.

11.6. An assessment of the argument-centred meta-method

There is no doubt that the argument-centred meta-method informs our philosophical (and perhaps scientific) knowledge. For example, Bigelow and Pargetter’s research tells us something important about the epistemology of acquaintance. Unfortunately, responding to qualia arguments one by one from the armchair can only tell us about very limited aspects of qualia. It can demonstrate why arguments against physicalism fail (if, in fact, they do so). In addition, using the example of Bigelow and Pargetter, it can further demonstrate the relevance of other philosophical fields, such as epistemology, to the study of qualia. However, developing a comprehensive understanding of qualia through this meta-method is unlikely.

Notably, this meta-method appears to be angled towards non-physicalist views on qualia. The non-physicalist might continue to formulate new arguments against physicalism or escape objections by pointing to some unexplained features (or rather unexplained assumed features) of qualia. It would
be difficult for the physicalist to reply to each argument one by one, unless a relatively comprehensive understanding of qualia is provided or shown to be forthcoming. However, this is not what arguments from the armchair are good at. Simply devising arguments from the armchair cannot tell us, for example, the empirical details about how the brain actually works. Unfortunately, this lack of empirical details about qualia might result in a lack of sufficient resources to reply to some arguments against physicalism. Obtaining such empirical details forces us to go beyond the argument-centred meta-method because philosophical argumentation ceases to be the centre of inquiry.

To be fair, though, the aims of researchers who adopt the argument-centred meta-method were typically not developing a comprehensive theory of qualia. While I believe that there is a need for the study of qualia to move beyond philosophical responses to arguments against physicalism, it is possible this advancement will be partly due to the philosophical knowledge provided by these researchers.

11.7. Advocating the event-centred meta-method

According to the event-centred meta-method, the main explananda for the study of qualia are the neurological events where qualia are taken to exist and our folk theory and introspection are produced as outputs. In this subsection, I advocate the event-centred meta-method because it allows us to revise our ontological assumptions and has a flexible methodology. These features of the event-centred meta-method allow us to avoid the inflexible assumptions present in both the theory-centred and the property-centred meta-methods, and hence allow us to, in some easier ways, develop scientific conceptions of qualia that can bridge the explanatory gap without disrespecting our phenomenal life and diverging from the historical path of science. In addition, it does not commit us to the narrowness of the argument-centred meta-method, as philosophical argumentation is not the only method available to interpret qualia events.

11.7.1. Revision of ontological assumptions

I argue that the event-centred meta-method allows for the revision of ontological assumptions because it holds no strong assumptions about the nature of qualia. As I have commented earlier, the theory-centred or the property-centred meta-method lead to a false dilemma between extremist affirmative and eliminative attitudes towards our assumptions about qualia, which has the potential to bias researchers. The event-centred meta-method, on the other hand, provides us with a way out of the dilemma. Remember, in the event-centred meta-method, qualia are observable events, phenomena
(scientifically speaking) that are tractable to a wide range of research programs; and, importantly, their hypotheses are revisable in light of discoveries. While assumptions are held, they are not inflexibly held and can be revised without being eliminative. What I believe to be the right choice of meta-method is one that can actually provide our current science with the greatest progress. Given that revised ontological assumptions could be useful grounds for future studies of consciousness to develop upon, I believe that it is one reason why the event-centred meta-method should be advocated.

It is important to note that my preference for the event-centred meta-method is not a claim that eliminativism is false. Even though the two are inconsistent – because the event-centred meta-method presupposes that there is an event, whereas the eliminativism insists that there are no such events – I acknowledge the latter as an open possibility. Following my discussion of revisions of old scientific concepts in section 11.5, there is no uncontroversial answer to what the connection between old scientific concepts and their newer counterparts is. In scientific practice, whether theoretical contents and natural kinds are revised (like <water>, <gravity>, or <acid>) or eliminated (like <caloric fluid>, <vital force>, or <demonic possession>) is most likely determined by pragmatic considerations of the scientists, and it is difficult to see why <qualia> should be an exception. In the future, there may be strong debunking evidence that can show that the seeming appearance of qualia is caused by, say, some neurophysiological processes too irrelevant or unworthy to be called qualia. However, before a criterion distinguishing eliminations from revisions is discovered and becomes uncontroversial among researchers, I believe that whether qualia should be eliminated should be determined by scientists (and philosophers) at that time according to their pragmatic concerns, not our armchair speculation.

11.7.2. Flexible methodology

We have seen how studies of qualia that both respect our phenomenal life and stay on the historical path of science are possible. One matter of concern is whether the scientific method can actually allow for the study of qualia. Consider Dennett’s heterophenomenology, which we discussed earlier. In Dennett’s view, it is the way that the sciences of consciousness like neuroscience and psychology actually work, and what he does is only spelling it out (1993b, p. 50, 2003, p. 20). Indeed, if it is true that the sciences of consciousness strictly require us to remain in the third-person perspective and to neglect the first-person perspective, then it is very difficult for the study of qualia and, indeed, our phenomenal life to be carried out. This is because qualia currently – and perhaps essentially – remain
a first-person experience and are thus difficult for purely third-person studies to investigate. If this is true, perhaps a study of qualia cannot be on the historical path of science and must diverge from it.

My approach to the worry will be as follows. I will show how the event-centred meta-method may lead us to accept an alternative approach to the sciences of consciousness, which I consider to be more attractive. Then I will show that the approach has already attracted a lot of followers and contributed to many influential researches in the scientific community.

No doubt, the event-centred meta-method, like the other meta-methods, represents a broad direction and attitude towards the study of qualia. Philosophers and scientists who share the same meta-method can utilise very different methodologies. For example, Dennett advocates methodological naturalism and Levine advocates philosophical mysticism regarding qualia, even though both adopt the property-centred meta-method. The event-centred meta-method is the same, but I argue that its openness to revisions of our ontological assumptions naturally, though not necessarily, motivates a wide and flexible methodology. The reason is that its theoretical revisability allows for methodological flexibility, which I will discuss in detail. I consider this a better foundation for further developments in the study of qualia, both scientific and philosophical, when compared with other contemporary methodological doctrines in the literature such as Dennett’s heterophenomenology.

We have seen that the event-centred meta-method allows ontological assumptions to be revised. It presupposes the existence of qualia but not any one description of them; and our initial theoretical conceptions about them are not simply assumed or required to be true, but rather simply allow us to begin our inquiry and is open to revision. If our theoretical – or ontological – assumptions are open to revision, then methodological assumptions should be the same because the two kinds of assumptions are interrelated. The precise reason is as follows. A methodological doctrine of a discipline is ultimately about how we can epistemically relate to the subject matter appropriately; hence, ontologies with regard to ourselves, the subject matter, and how the two can relate to each other always play some important roles in determining (part of) it. Accordingly, no one methodological doctrine can claim privileged or exclusive access to an event before empirical or philosophical evidence exists to support its claim.

A flexible methodology could be said to reflect a pragmatic scientific attitude towards the choice of methodological methods. With this attitude, all methods are considered hypothetical and experimental and – like the data collection methods in other areas of scientific research – have to be chosen with an open, pragmatic, permissive, and fallible attitude. As long as supporting evidence can
be provided, they are all accepted to help us to fix the target of investigation, to provide findings, and to prepare for the ideal final judgement in some cooperative ways, while every of them is considered fallible and challengeable by incoming evidence. The view is not anything novel but something that is widely accepted in other areas of science. Contrary to more dogmatic views of science that are often based on *a priori* considerations, I believe that the actual practice of science is flexible and ever changing. In principle, it permits all kinds of hypotheses, both theoretical and methodological, and cautiously asks for verification.\(^\text{62}\)

In the area of phenomenology, this naturally leads to a flexible phenomenology, i.e. a flexible methodology applied to phenomenology, reflecting the scientific attitude towards the choice of phenomenological methods. Possible methods that the flexible phenomenology can be open toward might include (not exhaustively) neuroscience, cognitive science, philosophy of mind, first-person reflection, heterophenomenology, and even the Cartesian, German, and Eastern first-person phenomenological traditions. This flexible phenomenology results in a resistance to the heterophenomenologist’s claim that her method has exclusive access through third-person scientific investigation. Heterophenomenology reflects Dennett’s ontological assumption, which is his eliminativism about phenomenological experience. This, I have argued, does not have to be assumed. Since the event-centred meta-method has no need to commit to Dennett’s eliminativism, it naturally resists his exclusive methodological assumption. Researchers adopting the meta-method do not have to stay in the third-person, and introspections can be taken as more than speech acts. Evidence does not have to be rejected simply because it is found within one’s first-person perspective.

 Nonetheless, it is important to note that while this flexible attitude resists exclusive assumptions, it does not reject them. Privileged or exclusive access is not considered as false, but is considered with an (cautiously) open attitude. It can be vindicated through empirical or philosophical evidence, even though current evidence for and against it should always be considered as fallible. In this case, heterophenomenology, for example, can still demonstrate its correctness through providing sufficient evidence.

\(^{62}\) Of course, some philosophers prefer to provide *a priori* conceptions of the so-called scientific method and thereby the distinction between science and pseudoscience or non-science are set aside. In my view, most areas of science have already developed their own ontological and methodological frameworks, and most of those frameworks are sufficiently sophisticated and systematic to allow appropriate researches to distant themselves from pseudoscience or non-science. Hence, there is no need to use *a priori* norms to provide such distinction. In fact, in my observation, it is currently more common for contemporary philosophers of science to observe what scientists actually do and offer philosophical regimentations, analyses, and advices. In contrast, it is not obviously an attractive approach for philosophers to develop an *a priori* set of rules that strictly restricts what scientists could do and ought to do.
evidence. What is important is that the evidence is not assumed to be authoritative but continuously assessed with the possibility of eventually being explained away.

In sum, with the flexible phenomenology, the methodological – much like the theoretical – should be flexibly adjusted by empirical considerations. There is no *a priori* methodological concern or norm that blocks scientific studies of qualia that bridge the explanatory gap without disrespecting our phenomenal life and diverging from the historical path of science.

One last point. Is the flexible phenomenology practical and useful to researchers? I think yes, and I believe that some researchers are already tacitly or explicitly applying it (or similar methodologies) and thereby have findings that are widely recognised as important breakthroughs. Dennett has claimed that ‘the ground rules [of heterophenomenology are] already tacitly endorsed by the leading researchers’ (1993b, p. 50), and that ‘most of the method is so obvious and uncontroversial that some scientists are baffled that [Dennett himself] would even call it a method’ (2003, p. 20). While it is true that scientists have typically shunned the inclusion of classical phenomenology, not all have. Recent research from neuroscience has shown a turn or, as neuroscientist Christof Koch notes, a ‘true paradigm shift’ from this tradition (2012, p. 6). For example, Masafumi Oizumi, Larissa Albantakis, and Giulio Tononi’s recent article ‘From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0’ (2014) provides an interesting and explicit example of scientists who take classical phenomenology of qualia seriously in order to understand its nature and then attempt to map this onto physical mechanisms. More specifically, they take some axioms in phenomenology to be ‘immediately evident’ and then map them onto their influential integrated theory (p. 2).

Despite this scientific research program proceeding in the complete opposite direction to what might typically be found in cognitive neuroscience, it has gained a wide amount of traction and followers.

### 11.8. Conclusion

In this section, I sketched a variety of meta-methods open to philosophers and scientists investigating qualia. Furthermore, I advocated for the event-centred meta-method as it allows for the revision of ontological assumptions and a flexible methodology. I argued that they are supported by the lessons

63 It can be questioned whether this method is associated with the event-centred metamethod because of its similarity to the property-centred metamethod. I think not because of the same line of reasoning I used to interpret Edelman and Tononi (2001) in Section 11.4.3, but this is not the concern here. What is important here is that the article is an obvious case to show some scientists’ turn from the tradition Dennett describes.
from the history and practice of science and, in addition, have the greatest likelihood to provide our current scientific (and philosophical) knowledge of qualia with the greatest progress. More importantly, it provides us with a way to bridge the explanatory gap without disrespecting our phenomenal life and diverging from the historical path.

Let us return to our major purpose. The major purposes of this section are to defend MN from the worry that qualia might be metaphysically non-naturalistic, and to defend CI from the worry that the nature of qualia might allow us to be acquainted with categorical properties. What I have defended is not a concrete theory of qualia but a meta-method for the study of them, and hence it remains an open question what qualia really are. What is crucial here is that there is no reason to expect that considerations of qualia will force us to either eliminate qualia and diverge from the historical path. Not to diverge from the historical path means being naturalistically respectable. In line with what we know about the historical path, it can also be expected that qualia, like other neurophysiological facts, are ‘ordinary matters’, to use J. J. C. Smart’s (1978, p. 340) expression; they should not be expected to be something extraordinary enough that can allow us to know of categorical properties in the way the categorical acquaintance thesis describes. In sum, we could have a reasonable expectation that qualia provide no threat to MN and CI.
This page intentionally left blank
Metaphysical naturalism and categorical ignorance
Part IV

Section 12. Metaphysical naturalism and categorical ignorance

12.1. Introduction

In Part I and II, I developed an account of the nature of metaphysical naturalism (MN) and the categorical ignorance thesis (CI) respectively. In Part III, I also discussed Russellian Monism (RM) and the view that categorical properties can be known through acquaintance with them in connection with the account and argued that they provide no formidable objection to the two doctrines that we are interested in.

Recall that the main purpose of this dissertation is to determine whether MN, a metaphysical framework shared by most contemporary analytic philosophers, can withstand the challenges provided by CI (if there are any). To be more precise, the following two problems will be addressed:

- **The problem of the possibility of natural categorical properties** (henceforth, the problem of possibility$_{NCP}$): Is it possible for there to be a property that is unknowable, categorical, and, at the same time, metaphysically natural?

- **The sceptical problem**: Does CI imply a scepticism about the metaphysical naturalness of categorical properties, which will, in turn, significantly decrease the probability that MN is true? (Recall that a default position of this dissertation is that the prior probability of MN, as shown by the traditional arguments for MN, is high [see the Introduction].)

In this final section, I will discuss these two problems.

Recall also that I have developed two conceptions of MN in Part I, which are the minimalist thesis and the metaphysical paradigm thesis. I will discuss the relation between each of these theses and the two problems respectively.

The structure of this section is as follows. In Section 12.2, I will introduce the problem of possibility$_{NCP}$ and the sceptical problem in detail. In Section 12.3, I will argue that the minimalist thesis is defensible against the two problems. In Section 12.4, I will argue that the metaphysical paradigm thesis is defensible against the two problems. In Section 12.5, I will conclude this section. In Section 12.6, I will conclude this whole dissertation.
12.2. The two problems

In this subsection, I will illustrate how the problem of possibility\(_{\text{NCP}}\) and the sceptical problem arise.

12.2.1. The problem of possibility\(_{\text{NCP}}\)

The problem of possibility\(_{\text{NCP}}\) concerns whether it is possible for there to be a property that is unknowable, categorical, and, at the same time, metaphysically natural. If the answer is no, any property cannot be both an unknowable categorical property and a metaphysically natural property; the simple fact that there are unknowable categorical properties, regardless of what they are like, is inconsistent with MN.

Some authors discuss an analogical question about the relation between unknowable categorical properties and physicalism instead of MN. The possibility question they concern is, unsurprisingly, the possibility of unknowable categorical properties that are physical. Those that are attracted to or lean toward a negative answer to the consistency question include Rae Langton (1998), David Braddon-Mitchell and Frank Jackson (2007), John Foster (1993), and Alyssa Ney (2007). Among these authors there are both (self-described) physicalists – Jackson, Braddon-Mitchell, and Ney – and (self-described) non-physicalists – Langton and Foster. The commonality between those authors is their belief that physicalism and the physical should be defined with reference to the posits of physical theories or what physics can have access to. Accordingly, because things that are unknowable, by definition, cannot be posited by physical theories, nor can they be accessible by physics, they cannot ever be physical. Hence, if unknowable categorical properties exist, then physicalism is false.

For example, as Langton puts it:

[T]hings as they are in themselves are not physical: for their intrinsic properties are not physical. Physics can discover nothing about things as they are in themselves. […] The physical world is a dependent world; things in themselves are independent. […] To say that the physical world is real, and that there is something nonphysical besides, is to deny physicalism. It is to deny that reality is exhausted by physical reality. (1998, pp. 207-208, original emphasis)

And as Braddon-Mitchell and Jackson put it:
[CI] concedes a lot to dualism. The core idea behind dualism is that the world exemplifies many properties over and above those mentioned in the physical sciences. [CI] agrees about this. (p. 141, original emphasis)

12.2.2. The sceptical problem

Different from the problem of possibility\textsubscript{NCP}, the sceptical question does not question the consistency between MN and the concept of unknowable categorical properties. On the contrary, it assumes that these two concepts are consistent. The concern of the problem is that regardless of whether MN is in fact true, if there are unknowable categorical properties, we in principle cannot know that they are metaphysically natural because we in principle cannot ever know of them. In particular, we might not justifiably reject alternative possibilities, such as Christian philosopher Foster’s (1993) idealism, according to which the categorical basis of everything is God’s mind. In sum, we in principle cannot know that MN is true, and thus can never be justified in believing in it.

A proponent of the sceptical problem is Sam Cowling, a critic of CI. He writes:

[I]f [CI] is true, we are ignorant of whether our world is one of objects and properties or a Tractarian world of facts or a Davidsonian world of events or even a Berkeleyian world of ideas in the mind of God. As should be clear, [CI] has troubling consequences for systematic metaphysics. (2010, p. 662)

In the following sections, I will discuss how the two conceptions of MN, which are the minimalist thesis and the metaphysical paradigm thesis, can respond to the two problems.

12.3. The minimalist thesis and CI

Recall that the minimalist thesis is as follows:

The minimalist thesis: Every entity or property instantiated in the actual world is metaphysically natural – (1) in the sense of (a) being posited by scientific concepts that are on the historical path of science or (b) being likely to appear on the path in the future, or (2) in the sense of playing only a set of theoretical roles that is closely similar to one played by a metaphysically natural entity or property in the former sense (see Section 2).
The historical path of science is roughly the process of modifying scientific concepts over the course of history, along which they are modified and finally developed into the versions in modern natural science; and the doctrine is minimalist in the sense that it requires its follower to avoid positing anything that is not adequately related to the historical path of science (in the way described by the above definition), such as God, demons, and souls.

In Sections 12.3.1 and 12.3.2, I will investigate how the minimalist thesis can respond to the two problems.

12.3.1. The minimalist thesis and the problem of possibility

To determine whether the problem of possibility would defeat the minimalist thesis, we might learn from the analogous debate concerning the conceptual consistency between physicalism and CI. The most important proposal to solve the apparent inconsistency between physicalism and CI in the literature is Daniel Stoljar’s use of the object-based conception of the physical, according to which the physical can be defined with reference to properties required to account for paradigmatic physical objects and those that supervene on these properties (see Sections 8 and 10 for a detailed introduction to and some discussions of the idea). Stoljar’s term ‘paradigmatic physical objects’, I have argued in Section 9.3.2, seems to mean objects we intuitively consider to be physical or objects most of us agree to be physical (or something very similar). Applying the same strategy to the minimalist thesis, paradigmatic metaphysically natural objects can be defined with reference to their adequate relation to the historical path, or their nature as bearers of properties that are adequately related to the historical path.

Unfortunately, as Langton and Christopher Robichaud argue, Stoljar’s object-based conception of the physical is too liberal (2010, p. 175). The worry is that it could count George Berkeley and Gottfried Leibniz as physicalists, because if, say, Berkeley’s view is true, God’s mind is required to account for paradigmatic physical objects. The same objection supposedly applies to a minimalist thesis that makes use of an object-based definition as well. Given that Stoljar’s object-based definition is not tenable, it seems that if we wish to defend the minimalist thesis from the problem of possibility.

---

64 An earlier version of the object-based conception of the physical is provided by Jackson, who calls the concept an ‘ostensive approach’ (1998, p. 7). But Jackson does not aim to use the conception to solve any problem similar to the problem of possibility.
possibility \textsubscript{NCP}, we need to develop an original proposal that can account for the conceptual consistency between the minimalist thesis and CI.\textsuperscript{65}

To develop such a proposal, it is important to recall the conception of unknowable categorical properties, which I expressed in Part II. In this view, it is possible that physical science can discover fundamental objects. For example, if electrons are fundamental objects, physics does not only posit a set of properties that we would attribute to an electron, but also posits \textit{the electron}. Unknowable categorical properties are believed to be those fundamental objects themselves or, if those fundamental objects are understood as theoretical roles, their players. We simply cannot know their exact nature because even though we can know of the scientifically knowable properties that they bear, we do not know exactly what their natures or features that realise those scientifically knowable properties are like.

Returning to the minimalist thesis, can those unknowable categorical properties be said to be adequately related to the historical path of science (as required by the definition of the minimalist thesis)? I believe the answer is yes, to a certain extent. The reason is that they can be understood as fundamental objects themselves or their role-players, and that physics in principle could posit fundamental objects. However, in a stricter and more demanding sense, the answer might be no. Given the assumption that natural science could only ever posit properties that the unknowable categorical properties bear, not their exact natures or features that realise those properties, it could still be said that since those categorical properties are not as current, future or ideal physics describes them. Hence, it could be said that they are not adequately related to the historical path. It is important to note that many scientists posit an unknowable side of physical objects that is more fundamental than the side described by scientific theories because they, much like theorists of CI, believe that natural science has limits to its reach. But it is still not very obvious that what they have in mind is exactly unknowable categorical properties, nor is it very obvious that the two kind of unknowable things play similar theoretical roles (see the definition of the minimalist thesis). Nevertheless, my own view is that it is still possible to count unknowable categorical properties as metaphysically natural if we consider the nature of the minimalist thesis carefully.

\textsuperscript{65} This very quick assessment of the object-based conception might not be fair enough, because both Stoljar and its earlier proponent, Jackson, admit that the conception has the weakness that is later pointed out by Langton and Robichaud (Stoljar 2001, p. 257n10; Jackson 1998, p. 7). Stoljar even makes it clear that one might supplement constraints to the conception in order to overcome the weakness (2001, p. 257n10). Hence, we should only understand the conception as one which is supposed to capture part of, not all of, the nature of physicalism. Nevertheless, because it is not my aim here to defend the conception, and also because of the limit of scope, I shall not discuss how the conception could be modified and whether its possible modifications are attractive.
My view is that there are two ways to understand the ‘minimal’. Understanding it as an all or nothing affair is only one of the two ways. Let me first illustrate the shortcoming of it before introducing the alternative way. If we only count those views that are consistent with the minimalist thesis in a very strict and demanding – and dogmatic – sense, then it is very difficult for there to be any discussion of the metaphysics of properties. For example, while the substratum theory posits a featureless substratum as the bearer of properties and the bundle theory posits an inexplicable relation that acts as a tie to bind the properties together, it is difficult to see how the two posits can be adequately related to the historical path. It can also be expected that many other areas of philosophy would face challenges if those who work on them wish to be naturalistically respectable. As Stoljar argues in his discussion of physicalism, it is difficult to see why we should accept a definition of physicalism that would count many respectable physicalists such as David Lewis, David Armstrong, and many others as non-physicalists when an alternative definition is available (see Section 8.3.2). The same can be said about MN.

The alternative version of the minimalist thesis I prefer is a moderate version. It understands the ‘minimal’ in terms of comparative degrees. According to this version, although some metaphysical views cannot be said to be minimal in a very demanding sense, they can be said to be naturalistically respectable because they are relatively minimal in the following sense. When their posits are added to a worldview that the demanding version of the minimalist thesis can respect, the worldview remains structurally similar and is not significantly more ontologically expensive. In the case of the substratum theory and the bundle theory, since their posits (i.e. featureless substrata and inexplicable relations) are some simple metaphysical explanations of our scientific theories, adding them to a naturalistically respectable worldview does not bring forth significant changes in its structure and ontological economy. Hence, the two theories could be counted as naturalistically respectable. In contrast, presumptively naturalistically unacceptable theories such as theism and panpsychism have posits like God or protophenomenal properties. Adding these to a naturalistically respectable worldview might bring forth more significant changes in its structure and ontological economy. This moderate version of the minimalist thesis seems to be more attractive, and preferable to the dogmatic way discussed above, because it can explain why most of us consider metaphysicians like Armstrong as metaphysical naturalists rather than non-naturalists.

Of course, the dividing line between the naturalistically respectable and unacceptable here is vague and has not been shown to be consistent. More work has to be done to sort it out; and this cannot be done here. Nevertheless, what is important for us is whether the moderate version of the
minimalist thesis can solve the problem of possibility\textsubscript{NCP}. I believe the answer is yes. This is because it seems that any reasonable dividing line between the naturalistically respectable and unacceptable should count my view on categorical properties as naturalistically respectable. Let me illustrate. According to the moderate version of the minimalist thesis, whether a view is naturalistically respectable or not is a comparative matter. In comparison with most kinds of things that are not adequately related to the historical path of science, it is obvious that the changes that adding unknowable categorical property would make to a naturalistically respectable worldview are insignificant. This is because of the strong relation they bear to the scientifically observable fundamental objects. Recall that in my view they are those objects or their role-players, and their features of natures are nothing more than the realisers of the scientifically observable properties those objects bear. With this in mind, it can be said that unknowable categorical properties and scientifically observable fundamental objects are not very distinct. Therefore, even though the dividing line between the naturalistically respectable and unacceptable here is vague, this should not be an issue for the naturalistic respectability of my view on unknowable categorical properties because it is presumably not close to any reasonable dividing line.

12.3.2. The minimalist thesis and the sceptical problem

Following the discussion in Section 12.3.1, it is possible that unknowable categorical properties are metaphysically natural in the sense that a moderate version of the minimalist thesis requires. But the sceptical problem remains a challenge. Applying the problem to our current picture, the worry can be stated as follows. If CI is true, we cannot really be sure that categorical properties do not have extra features or natures that cannot be categorised as metaphysically natural, even according to the moderate version of the minimalist thesis. For example, it is possible that categorical properties are a single cosmic mentality as cosmopsychism describes. Since we cannot ever know of the categorical properties, we cannot justifiably deny such a possibility.

It is worth noting that Cowling is wrong in believing that if the sceptical problem (even in its strongest form) is not resolved, then there are absolutely no disagreements on the nature of categorical properties that we might be able to settle. Consider one such disagreement: whether Berkeleyan idealism is true or not. We might gain some traction because Berkeleyan idealism does not only posit categorical properties, but also posits causal behaviours and dispositions of things. For example, Berkeleyan idealism explains physical events by the hypothesis that our minds are directly receiving sensory information from God’s mind. Moreover, George Berkeley himself, as a Christian theist,
believes that God is a person who could perform, and indeed has performed, miraculous actions such as the resurrection of Jesus. The sceptical problem does not challenge our judgments on causal and dispositional matters like such cases. Therefore, it is possible that the metaphysical naturalist could justifiably reject Berkeleyan idealism solely on the basis of considerations of the causal and dispositional structure of the world. Nevertheless, I will set this issue aside, because not all disagreements on the naturalistic respectability of categorical properties can be settled in this way.

I believe that the sceptical problem can be responded to in the following way. Philosophers like Cowling, who believe that CI is the thesis that we could not ever have knowledge of any kind regarding categorical properties, have simply made a mistake. CI is the thesis that we cannot know what categorical bases are precisely like, not that we cannot attribute abstract metaphysical categories such as being self-identical, being properties, and being unknowable (as CI describes) to those properties (see Section 7.2). For example, if we accept that it is metaphysically necessary that everything is self-identical, then we can, undoubtedly, know of that fact that categorical properties (if there are any) are self-identical as well.

In line with what has been said, we may also know of, or have a reasonable ground to believe in, categorical properties’ belonging to some other abstract metaphysical categories, such as being metaphysically natural, by inference to the best explanation. To say that the moderate version of the minimalist thesis is wrong about unknowable categorical properties is to say that other than the natures or features that realise the scientifically observable properties they bear, they have the additional nature of, say, being protoconsciousness or a single cosmic mentality. As Graham Oppy notes, we could understand MN and its major rivals like theism in a sense where MN’s rivals are the ‘extensions’ of it – i.e. the rival views have all the naturalistic commitments that MN has, but add additional naturalistically unrespectable commitments (2013, p. 88). He then argues that provided all other factors (such as philosophical arguments, explanatory powers, and so on) are equal, MN performs better than its rivals in terms of ontological economy because it is simpler. Applying Oppy’s argument to our discussion of categorical properties, it seems that the ‘best explanation’ of categorical properties is that they are metaphysically natural (in the way the moderate version of the minimalist thesis defines it), in the case that we are not convinced by arguments for the metaphysically non-natural, such as qualia arguments (see Sections 8-11). This provides us with reasonable grounds to believe that MN is true, even if unknowable categorical properties are taken into account.

No doubt, the ‘best explanation’ is fallible and not certain, and has its own set of probabilities to be true or false. Hence, it presumably cannot allow the prior probability of MN to persist but will,
instead, lead to a change in the probability that MN is true. Nevertheless, it is not obvious that the change in probability caused is significant, let alone whether it is an increase or a decrease. In sum, it is not obvious that the sceptical problem has sufficient force to defeat the minimalist thesis.

12.4. The metaphysical paradigm thesis and CI

Recall that the metaphysical paradigm thesis is as follows:

**The metaphysical paradigm thesis:** Every entity or property instantiated in the actual world is metaphysically natural – in the sense of being compatible with the metaphysical paradigm shared by the core theories (see Section 2).

Core theories are theories that have most if not all of their main concepts on the historical path; and these core theories share some metaphysical commitments, which I take to be a metaphysical paradigm I call the core paradigm (CP). The metaphysical paradigm thesis is the view that MN is CP, and that its contents are the metaphysical commitments of CP. In Section 3, I offered a list of the metaphysical commitments. The list is not argued to be historically accurate because this dissertation is not a historical study. Instead, it is an example of how CP can be understood intuitively. It is as follows:

1. Everything is constituted by (something similar to) theoretical entities in physics.
2. Nothing is divine.
3. There are no fundamental and nonlocal mental properties.
4. There are no vital forces.
5. There are no fundamental teleological properties.
6. Everything can only be empirically known if it can be known.
7. Everything operates according to laws of nature.
8. There is a uniformity of everything.

In Sections 12.4.1 and 12.4.2, I will discuss how the minimalist thesis can respond to the two problems.
12.4.1. The metaphysical paradigm thesis and the problem of possibility

It seems to be obvious that half of the commitments of CP are consistent with CI. For example, take commitment (2) – according to which nothing is divine. It is possible that categorical properties are not divine, and hence the fact that they exist is not conceptually incompatible with the commitment. The same can be said about the following commitments (3), (4), and (5). What we are left with are commitments (1), (6), (7), and (8), and I will discuss them respectively in the following paragraphs.

Commitment (1) – according to which everything is constituted by (something similar to) theoretical entities in physics – does not seem to be conceptually inconsistent with CI as well. It is true that categorical properties are not constituted by theoretical entities in physics, because they are their realisers. But it is noteworthy that we must consider self-constitution as a kind of constitution when we understand this commitment, because there is no reason to exclude theoretical entities in physics that are not constituted by other theoretical entities in physics. Since fundamental objects can be expected to be theoretical entities in future physics, and categorical properties, in my view, are those objects or the role-players of them, it seems that categorical properties can be counted as things that are constituted by (something similar to) theoretical entities in physics.

Commitment (6) – according to which everything can only be empirically known if it can be known, again, does not seem to be conceptually inconsistent with CI. It seems to be obvious that we, in fact, know of their existence by empirical means. Of course, we cannot directly empirically observe categorical properties; instead, we come to know of their existence by philosophical reflections on what the world is like, which, in turn, is known by direct empirical observations and other philosophical reflections on them. Nevertheless, most contemporary philosophers would consider facts that are known by reflections on empirically observations as empirically known.

Commitment (7) – according to which everything operates according to laws of nature – is also a commitment that seems to be consistent with CI. Of course, categorical properties are supposed to be bearers of causal and dispositional properties, not the causal and dispositional properties they bear. Because the term ‘operate’ means causal behaviours, to a certain extent categorical properties might be said to be unable to operate. Nevertheless, if the causal and dispositional properties they bear operate according to laws of nature, it is difficult to see why we cannot describe them as operating according to laws of nature as well.

Commitment (8) – according to which there is a uniformity of everything – is perhaps the only case in which the consistency with CI is not very obvious. In Section 3.3.2.18, I explained that by
‘uniformity’ I mean only one theoretical framework is required to describe all entities or properties, with the exception of those whose constituents can be described by the theoretical framework. This seems to be inconsistent with CI because categorical properties, according to this thesis, cannot be (precisely) described by any theoretical framework. As Braddon-Mitchell and Jackson note, ‘The core idea behind dualism is that the world exemplifies many properties over and above those mentioned in the physical sciences. [CI] agrees about this’ (2007, p. 141, original emphasis). Nevertheless, since categorical properties realise the scientifically observable properties they bear, the latter can be said to be a manifestation of the former. It could be said that the descriptions provided by our theoretical framework – mass, size, and charge – are the best theoretical representations of the categorical properties, even though they cannot precisely describe them, telling us exactly what they are like. Hence, it seems that to a certain extent it can still be said that only one theoretical framework is required to describe all entities or properties.

The above way of handling the relation between unknowable categorical properties and commitment (8) might not be compelling to everyone. This is because even if a theoretical framework can provide a best theoretical representation of categorical properties, it is presumably not one that can apply perfectly. Hence, it seems to be an obvious fact that categorical properties are not perfectly uniform with other things. Nevertheless, there is another perspective that we could use to assess the issue. In Section 3.3.2.18, we have seen the descriptions of the doctrine of uniformity made by several philosophers. For example, Braddon-Mitchell and Jackson note that the core idea of the doctrine is that ‘nature is uniform within and without the mind’, as opposed to the dualist view that ‘things in the mind are very different from things in the rest of the world’ (2007, p. 33). Gregg Rosenberg notes that the doctrine is ‘to place human beings in the world without making special, ad hoc assumptions that are discontinuous with everything else we have good reason to believe about nature’ (2004, p. 8). From both of these remarks, we see that the doctrine that weighs most is inter-domain uniformity. The idea is that there is always a uniformity between the inside and outside of a particular domain, such as the human mind. This is different from intra-domain uniformity, which is a uniformity within a particular domain. Returning to the concern of categorical properties, it is presumably universal in the actual world that causal and dispositional properties have their categorical bases. Hence, even if there is a disuniformity between causal and dispositional properties and their categorical bases, it is not inter-domain but merely intra-domain. With this in mind, it is not the kind of disuniformity that commitment (8) is most resistant to. Therefore, it could be said that even though commitment (8) is not perfectly consistent with CI, there can be an acceptable accommodation between the two doctrines.
I have now shown that all of the commitments of CP (I spelled out) are consistent with CI. Hence, it can be concluded that the metaphysical paradigm thesis is consistent with CI. Therefore, it is possible for there to be unknowable categorical properties that are metaphysically natural in the sense defined by the metaphysical paradigm thesis.

12.4.2. The metaphysical paradigm thesis and the sceptical problem

In Section 12.4.1, I have shown that it is possible for there to be unknowable categorical properties that are metaphysically natural in the sense defined by the metaphysical paradigm thesis. The sceptical problem, in the case of the metaphysical paradigm thesis, is then whether unknowable categorical properties in the actual world can be known to metaphysically natural, in the sense that all of the metaphysical commitments are true about them.

In Section 12.3.2, when I discussed the relation between the minimalist thesis and the sceptical problem, I handled the sceptical problem by inference to the best explanation. The strategy was to argue that for the minimalist thesis to be wrong about unknowable categorical properties, those properties have to have some additional metaphysically non-natural features or natures other than those that realise the scientifically observable properties they bear. Provided that all other factors (such as philosophical arguments, explanatory powers, and so on) are equal, a simpler explanation should be considered as a better explanation, and hence we have a reasonable ground to believe in the minimalist thesis because it does not have the additional commitments stated above.

I believe that the same is true about the relation between the metaphysical paradigm thesis and the sceptical problem. I have shown in Section 12.4.1 that if unknowable categorical properties only have features or natures that are required to realise scientifically observable properties, there is nothing inconsistent with the metaphysical paradigm thesis. In order for the metaphysical paradigm thesis to be wrong about unknowable categorical properties, those properties have to have additional features or natures that are metaphysically non-natural (in the sense that their existence violates some metaphysical commitments of CP). But provided that all other factors are equal, a simpler explanation should be considered as a better explanation, and hence we have a reasonable ground to believe in the metaphysical paradigm thesis because it has less commitments.

12.5. Conclusion of this section

I have shown that both versions of MN (i.e. the minimalist thesis and the metaphysical paradigm thesis) are defensible against the challenges of the problem of possibility_{NCP} and the sceptical problem.
that arises in the consideration of CI. Both theses are consistent with CI and can be justifiably believed in the presence of CI. No doubt, unknowable categorical properties are mysterious. Their existence increases the uncertainty of what the world is like and thus might decrease the probability that MN is true. But there is no evidence that the decrease in probability (if there is any) is significant. In sum, CI does not seem to be a significant threat to MN, as opposed to what some authors like Langton, Foster, Braddon-Mitchell and Jackson, Ney, and Cowling might believe or worry about.

12.6. Conclusion of this dissertation

The purpose of this dissertation is to determine whether metaphysical naturalism (MN), which is a metaphysical doctrine or framework that many contemporary philosophers align themselves with or tacitly assume, will be challenged by the categorical ignorance thesis (CI). In Parts I and II, I defended an account of the nature of metaphysical naturalism (MN) and the categorical ignorance thesis (CI) respectively. In Part III, I also discussed Russellian Monism (RM) and the view that categorical properties can be known through acquaintance with them, and argued that they provide no formidable objection to the two doctrines that we are interested in. In this final part and final section, making use of the views I developed in Part I and II, I argued that MN and CI are consistent, and that MN can be justifiably believed in the face of the possible decrease of its probability caused by CI.
References


Dehaene, S 2014, Consciousness and the brain: deciphering how the brain codes our thoughts, Viking, New York.


Heil, J 2003, From an Ontological Point of View, Oxford University Press, Oxford.


253


Quine, WVO 1995, ‘Naturalism; or, living within one’s means’, Dialectica, vol. 49, pp. 251-263.


 Schroer, Robert 2010a, ‘How far can the physical sciences reach?’, American Philosophical Quarterly, vol. 47, no. 3, pp. 253-266.


Strawson, G 2008a, ‘The identity of the categorical and the dispositional’, *Analysis*, vol. 68, no. 4, pp. 271-282.


Appendix A. The meaning(s) of the term ‘naturalism’

In Part I of this dissertation, I provided an account of what metaphysical naturalism is. But the term ‘naturalism’ has very diverse – and sometimes inconsistent – meanings in both philosophical and folk literature; and metaphysical naturalism is only one of the many doctrines that the term might stand for. As David Papineau writes, ‘Many different ideas parade under the banner of philosophical naturalism’ (2014, p. 116). In this dissertation, I focused exclusively on and discussed metaphysical naturalism as if it is an independent doctrine, but not everyone would agree with this way of understanding it. With the aim that the possible disagreement can be resolved, this appendix is a survey of the doctrines that the term ‘naturalism’ might stand for, with a discussion of why I believe they should be considered as independent of each other.

There are least six different, sometimes incompatible, doctrines (and also their variants) that the term ‘naturalism’ might stand for. They include what I call (1) ‘metaphysical naturalism’, (2) ‘epistemological naturalism (in accordance with metaphysical naturalism) (ENAMN)’, (3) ‘methodological naturalism (as the term is typically used in philosophy)’, (4) ‘naturalisation projects’, (5) ‘methodological naturalism (as the term is typically used in science)’, and (6) ‘indispensability naturalism’.

A.1. Metaphysical naturalism

Metaphysical naturalism (also called ontological naturalism and, controversially, ‘antisupernaturalism’) is a doctrine held by a number of philosophers and the folk. According to the doctrine, conceivable entities, properties, relations, and states of affairs can be categorised as being ‘metaphysically natural’ or ‘metaphysically non-natural’ (also, controversially, called ‘supernatural’), and the actual world contains only the ‘metaphysically natural’ ones. Since the doctrine has already been discussed in detail in the Introduction and Part I, I shall not go any further here.

A.2. Epistemological naturalism (in accordance with metaphysical naturalism) (ENAMN)

‘Epistemological naturalism (in accordance with metaphysical naturalism) (ENAMN)’ is an epistemological implication of metaphysical naturalism. It should not be confused with ‘methodological naturalism (as the term is typically used in philosophy)’ and ‘naturalised epistemology’ (both of which I will introduce in this appendix), even though these three distinct philosophical views are all often called ‘epistemological naturalism’ in the literature. According to
ENAMN, the knowing faculties and process of human being, including sensations, perceptions, cognitions, the relation between the knower and the known, and so forth, are all metaphysically natural. What follows is that we only have epistemic access to things that can be known through metaphysically natural knowing faculties and process; any belief that cannot be known in this way must be unjustified. With Evan Fales (2007, p. 126) as an exception, not many philosophers have explicitly used the term ‘epistemological naturalism’ to refer to the doctrine, but its content is often expressed in the works of influential writers on metaphysical naturalism, such as David Armstrong (1978) and Andrew Melnyk (2009), as a methodological implication of metaphysical naturalism. Other authors often speak of a methodological implication of physicalism in the same sense. Examples include what Hartry Field calls ‘the methodological role of physicalism’ (1992, p. 271) and what Carl Gillett calls ‘methodological physicalism’ (2001, p. 225). Given that physicalism is a version of metaphysical naturalism, these doctrines must be versions of ENAMN.

Notable examples of the use of the doctrine include Sharon Street’s (2006) Darwinian dilemma against moral realism and Paul Benacerraf (1973) and Hartry Field’s (1989) Benacerraf-Field challenge against Platonism in philosophy of mathematics, even though they do not explicitly use the term ‘naturalism’. According to Street’s Darwinian dilemma, it is unclear how moral properties can have causal influences on the evolution of human ‘evaluative attitudes’ (2006, p. 109). Moral realism is therefore unjustified. According to the Benacerraf-Field challenge, if Platonist mathematical entities (i.e. numbers) exist objectively in the mind-independent world, ‘there seems to be a difficulty in seeing how we could ever know that they exist, or know anything about them’ (Field 1989, p. 25). Mathematical Platonism is therefore unjustified.

A.3. Methodological naturalism (as the term is typically used in philosophy)

Methodological naturalism (as the term is typically used in philosophy), or one of the three understandings of the term ‘epistemological naturalism’, is a different doctrine from EMAMN. It is a methodological exclusivism, according to which the scientific method is the only justified method to acquire knowledge – or, more moderately, the only justified method of doing philosophy and/or perhaps other academic works. The discipline of philosophy should thereby operate in a way that is an extension of the discipline of science, rather than a ‘first philosophy’ that is prior to science (and other knowledge). Because of its emphasis on the dominant and exceptional role of science, the doctrine is sometimes called ‘scientism’, even though methodological naturalists rarely use this term as a self-characterisation (Keil 2008, p. 266).
The doctrine was popularised by W. V. O. Quine, who is widely recognised as the most distinguished representative of the term ‘naturalism’ in the 20th century, even though the doctrine can be traced back to some American philosophers in the first half of the 20th century, such as John Dewey, Sidney Hook, Ernest Nagel, George Boas, Abraham Edel, Sterling Lamprecht, and John Herman Randall (Kim 2003, p. 86; Papineau 2015). In Quine’s view, there is no methodological standpoint external to science. The role of philosophy is just ‘to clarify, organize, and simplify the broadest and most basic concepts, and to analyse scientific method and evidence within the framework of science itself’ (1995, p. 257). As he puts it:

My position is a naturalistic one; I see philosophy not as an a priori propaedeutic or groundwork for science, but as continuous with science. I see philosophy and science as in the same boat—a boat which, to revert to Neurath’s figure as I so often do, we can rebuild only at sea while staying afloat in it. There is no external vantage point, no first philosophy. (1969, p. 126)

Quine’s methodological naturalism is based on his denial of a priori knowledge, that is knowledge independent of experience (even though every available definition of the a priori is controversial), because it is often seen as the domain that distinguishes philosophy from other disciplines (including science). His most influential contribution to this goal is his criticism of the analytic/synthetic distinction in his seminal article ‘Two dogmas of empiricism’ (1951), which he believes results in ‘a blurring of the supposed boundary between speculative metaphysics and natural science’ (1951, p. 20). In the article, he argues that the analytic/synthetic distinction in philosophy, which is roughly the distinction between linguistic truths that are factually neutral and factual truths, is flawed, because all linguistic truths are, in fact, grounded in experience. The analytic, which is commonly considered the major component of the a priori, should therefore be rejected. With this in mind, he considers every statement, including logical laws, to be open to scientific revision:

[N]o statement is immune to revision. Revision even of the logical law of the excluded middle has been proposed as a means of simplifying quantum mechanics; and what difference is there in principle between such a shift and the shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle? (1951, p. 40)
Quine’s view has attracted numerous influential methodological naturalist followers such as Michael Devitt (1998). Frank Jackson observes, ‘The Quinean position […] was seen as radical when it was first propounded. Nowadays it is close to orthodoxy’ (1998, p. 52).

Nevertheless, Quine’s denial of the a priori and the unique role of philosophy remains highly controversial even among methodological naturalists. Methodological naturalism is an exclusive methodological commitment to scientific method, and the Quinean view is not the only attempt to make this possible. Some methodological naturalists have proposed ‘harmless’ conceptions of the a priori that they consider to be compatible with considering scientific method as the only justified method (see, for example, Kitcher 1980; Field 1996; Rey 1998). Others consider the a priori to be indispensable but philosophically insignificant (see, for example, Papineau 2011). There are, furthermore, attempts to reconcile the a priori and methodological naturalism by assigning the a priori a limited role, which is simply to analyse the definitions of folk terms, while the job of inquiring into ontology is completely left to science (see, for example, Braddon-Mitchell 2009). In sum, while methodological naturalism is an exclusive methodological commitment to scientific method, how this methodological commitment should be applied to philosophy remains an open question.

A.4. Naturalisation projects

‘Naturalisation projects’ (e.g. naturalisations of mental phenomenon, moral naturalism, and a third understanding of the term ‘epistemological naturalism’, also called naturalised epistemology) are attempts to explain the existence of, or replace in our ontology, particular apparently ‘non-natural’ entities or properties (e.g. mental properties, moral norms, and epistemic norms) by some ‘natural’ entities or properties (e.g. brain states, pain, pleasure, and those entities or properties found in cognitive psychology). An example is a utilitarian moral naturalist’s identification of good with pleasure.

Even though the concepts of naturalness and non-naturalness concerned in these naturalisation projects are sometimes the same ones dealt with by metaphysical naturalism, this is often not the case. What these terms actually predicate depends on what the philosopher intends to mean when she uses them. For example, in the context of metaethics, the terms ‘naturalness’ and ‘non-naturalness’ stand for multiple meanings. While their meanings can be the ones prescribed by metaphysical naturalism (see, for example, Wedgwood 2007, p. 197), they often, instead, predicate the two sides of the fact/value distinction. In the latter case, ‘naturalness’ is used to refer non-normative facts, as opposed to norms (see, for example, Railton 1986, p. 166); the term ‘moral naturalism’ merely predicates a
reductionist (or supervenient) view of value, according to which moral (and other) norms can be
reduced to non-normative properties. Some others consider the terms ‘naturalness’ and ‘non-
naturalness’ to refer to exclusive empirical knowability and a priori knowability (Copp 2003, p. 185).
Just as the term ‘naturalism’ has varied meanings, so too do the terms ‘natural’ and ‘non-natural’ as
used by those who pursue naturalisation projects.

The motivations of naturalisation projects are often metaphysical naturalism (and ENAMN) and
methodological naturalism (even in cases where ‘naturalness’ and ‘non-naturalness’ do not stand for
the meanings prescribed by metaphysical naturalism) but are not necessarily so, since there are some
metaphysical non-naturalists and methodological non-naturalists that are interested in them. For
example, a Christian physicalist believes in the naturalisation of the human mind while still holding
beliefs in religious epistemology, the existence of a divinity and physical resurrection (see, for
example, Corcoran 2006). This is because her belief in naturalism is local but not universal; her
naturalism is limited to the area of the human mind but it does not extend to other things, such as the
existence of God and miracles. Therefore, the view is sometimes also called local naturalism as
opposed to global naturalism.

Although the success of a naturalisation project does not imply metaphysical naturalism or
methodological naturalism, naturalisation projects are often important topics for metaphysical
naturalists and methodological naturalists because of their need to explain the apparently non-
natural properties. However, it is, again, noteworthy that it is not always the case. As David Copp says, ‘Moral
naturalists and non-naturalists are moral realists. They agree that there are moral properties’ (2012, p.
24). In terms of attitudes towards specific kinds of apparently non-natural properties, such as moral
properties, not every metaphysical naturalist philosopher or methodological naturalist philosopher is
a realist. An alternative approach for metaphysical naturalists and methodological naturalists is
antirealism or eliminativism, which is to suggest that those apparently non-natural properties do not
really exist but are useful fictions, incomplete understandings, or even illusions. Classical examples
include J. L. Mackie’s (1990) moral error theory and Paul Churchland (1981), Patricia Churchland
(1988), and Daniel Dennett’s (1991, 1993a) eliminative materialism about (some) mental phenomena.
In these cases, those apparently non-natural properties can still be recognised as non-natural, since
they are not naturalised, but eliminated.

Naturalised epistemology and methodological naturalism are both often called ‘epistemological
naturalism’, and both are strongly influenced by Quine. In fact, the term ‘naturalised epistemology’
originates from Quine’s seminal article ‘Epistemology naturalized’ (1969). But the two concepts are
separate and should not be confused. Naturalised epistemology is (roughly) the attempt to explain the existence of (or replace in our ontology) epistemic norms by some natural properties. (See, for an example of a replacement thesis, Quine 1969. See, for an example of a supervenience thesis, Kim 1988.) It is not the doctrine about general philosophical methodology that I introduced earlier in Appendix A.3

A.5. Methodological naturalism (as the term is typically used in science)

Methodological naturalism (as the term is typically used in science) is entirely different from methodological naturalism (as the term is typically used in philosophy). It is a methodological view about scientific methods but not philosophical methods and is often seen in debates over the separation of science and religion, especially debates over the legitimacy of creationism as a scientific theory. In this view, science ought to be ideologically and religiously neutral and therefore ought to take metaphysical naturalism as its assumption. When scientists attempt to explain natural phenomena, they should not appeal to any supernatural entities, especially divinity, even if they personally believe in their existence. A modest version of the doctrine is that appeals to supernatural entities should not be ruled out in principle but should be taken as ‘a last resort’ (Draper 2005, p. 298; Fales 2007, p. 123). What follows is that believers and non-believers of religion practice science in the same way because it is a domain free of religious commitment (Papineau 2015).

It should be noted that the metaphysically naturalistic assumption held by the doctrine is merely a strategic assumption but not a truth claim. Methodological naturalism remains neutral about whether metaphysical naturalism is actually true. This means that the methodological naturalist assumes metaphysical naturalism as though it is true only when she does science. She can be a metaphysical naturalist or metaphysical non-naturalist because of her own religious or philosophical commitment, or because of her personal experience.

Due to the popularity and influence of a seminal article on the isolation of science and religion by the famous evolutionary biologist Stephen Jay Gould, ‘Nonoverlapping magisteria’ (2001), methodological naturalism is also sometimes considered to be the doctrine that science is not taking a stand on the ‘ultimate questions’ (Ruse 2010, p. 236), such as morality and the existence of a necessary being.

A.6. Indispensability naturalism

‘Indispensability naturalism’ is an original term I use to describe a doctrine that is currently relatively unpopular. P. F. Strawson calls it ‘soft naturalism’ and ‘liberal naturalism’ (1985, p. 1); Barry Stroud calls it ‘open-minded naturalism’ and ‘expansive naturalism’ (2004, p. 34). According to the doctrine, we have an inevitable commitment to some common-sense beliefs for the reason that they are necessitated or implied by our human nature, intuitions, or real life experience that we cannot possibly avoid having, and hence we are justified in believing in them. As Strawson puts it, ‘it recognizes the human inescapability and metaphysical acceptability of […] various types of conceptions of reality’ (1985, p. 68).

The most notable and seminal contemporary contributors to this doctrine include P. F. Strawson and Barry Stroud, both of whom attribute the origin of this doctrine to the 18th century Scottish philosopher David Hume. As Strawson explains the doctrine in his seminal book Scepticism and naturalism: some varieties (1985):

> [O]ur natural disposition to belief, on the points challenged by the sceptic, is absolutely compelling and inescapable; neither shaken by sceptical argument nor reinforced by rational counter-argument. Where Nature thus determines us, we have an original non-rational commitment which sets the bounds within which, or the stage upon which, reason can effectively operate, and within which the question of the rationality or irrationality, justification or lack of justification, of this or that particular judgment or belief can come up. (1985, p. 39)

And as Stroud explains the doctrine in his article ‘The charm of naturalism’ (2004):

> What I am calling more open-minded or expansive naturalism says we must accept everything we find ourselves committed to in accounting for everything that we agree is so and want to explain. We want to explain the thoughts, beliefs, knowledge, and evaluative attitudes that we think people have got. (2004, p. 34)

Inevitable commitments indispensability naturalists have in mind include those regarding the existence of causation, the existence of an external world, the immaterial nature of mind, and so forth. Some alternative views such as radical scepticism and reductive physicalism are rejected by their
considerations. With this in mind, the ‘natural’ in ‘natural-ism’ in this context seems to refer to our ‘human nature’ or a ‘natural inclination’ to hold some intuitive beliefs.

A.7. A variant: liberal naturalism

I have listed the six different, sometimes incompatible, doctrines that the term ‘naturalism’ most often stands for. The six doctrines each have their variants, but I cannot possibly hope to introduce them all. However, I will discuss one influential variant: what Mario De Caro, Alberto Voltolini, and David Macarthur call ‘liberal naturalism’ (De Caro & Voltolini 2010; De Caro & Macarthur 2010). I do not consider it to be an independent kind of naturalism like the above six doctrines because it is a variant of metaphysical naturalism.

In terms of methodology, liberal naturalism rejects the monist methodological notion of methodological naturalism that accepts only the scientific method. It advocates a methodological pluralism, according to which there are justified methodologies other than the scientific method. In terms of ontology, it does not require that all metaphysically non-natural entities are either reduced or eliminated. According to De Caro and Voltolini, examples of metaphysically non-natural entities being accepted include reasoning, mathematical entities, normativity, intentionality, modality, responsibility, freedom, and consciousness (2010, p. 70). But liberal naturalism is still claimed to be a kind of naturalism, because all of the metaphysically non-natural entities above, according to the liberal naturalist, are compatible with the casual closure of the metaphysically natural, which is the idea that there are no interventions of the causal order of the metaphysically natural by the metaphysically non-natural.

De Caro, Voltolini, and Macarthur attribute the doctrine to a group of philosophers, which includes several influential philosophers such as John McDowell, Hilary Putnam, Huw Price, Akeel Bilgerama, and John Dupré (De Caro & Voltolini 2010, p. 77; De Caro & Macarthur 2010, p. 10-11). But I merely attribute liberal naturalism being introduced here to De Caro, Voltolini, and Macarthur, and leave it as an open question whether the attribution of the doctrine to other philosophers by the three writers is correct. This is because whether the view of each of the philosophers really possesses the features of liberal naturalism described above is a question in need of further assessment, which I cannot undertake here.
A.8. John McDowell’s naturalism

Among the views of philosophers to which De Caro, Voltolini, and Macarthur have attributed the term ‘liberal naturalism’, the most influential is (arguably) that of John McDowell, which seems to be the origin of the term ‘liberal naturalism’ (see McDowell 2004, p. 95) and is often considered a significant contribution to the contemporary literature on naturalism. Here I will introduce McDowell’s view in detail because of its seminality and importance in the literature on naturalism.

Following Wilfrid Sellars, McDowell argues that there is a ‘space of reasons’ apart from the ‘realm of law’. The realm of law is composed of causal laws of nature, which we know through natural science, while the space of reasons is composed of autonomous and irreducible thinking and reasoning (1994, p. 71-76; 2004, p. 93-94). What realises the space of reasons is the ‘second nature’ of human being, which is our ‘practical wisdom’, such as our conceptual and norm-following capacities (1994, p. 84). In McDowell’s view, the space of reasons and our second nature are not natural phenomena (2004, p. 92), but they are developed – through processes such as ethical upbringing – upon the basis of the ‘potentialities that belong to a normal human organism’ (1994, p. 84), which are natural. He rejects what he calls ‘bald naturalism’ or ‘restrictive naturalism’, which attempts to reduce the space of reasons and the second nature into the realm of law, since he, instead, considers the space of reasons and our second nature to be irreducible and thus sui generis (2004, pp. 93-94), i.e. of its own kind.

What is worthy of consideration here is that McDowell’s ‘liberal naturalism’ is a view that does not obviously possess the features of the ‘liberal naturalism’ advocated by De Caro and Voltolini. While the latter advocates the view that there is something ‘ontologically independent’ about the metaphysically natural, this does not seem to be what McDowell’s liberal naturalism advocates. In McDowell’s view, even though the space of reasons and our second nature are not categorised as being natural, they – in his words – have ‘a foothold in’ and ‘could not float free of’ laws of nature and the biological, which are natural (1994, p. 84). As I understand this, this is a supervenience relation in the standard terminology of contemporary philosophy of mind, which is an ontologically dependent relation rather than an ontologically independent one.

In fact, it is difficult to see why McDowell’s liberal naturalism should be differentiated from standard metaphysical naturalism, since the doctrine shares a similarity with a popular view among metaphysical naturalists and even physicalists, which is the view that special science laws (such as economic laws, psychological laws, and biological laws) supervene on physical laws but are not
Reducible to them (see, for example, Fodor 1974; Papineau 2001; Loewer 2001). Different explanations are provided to explain the irreducibility of special sciences. One common explanation is that physics is the study of global laws, while special sciences are the study of non-global laws (see, for example, Loewer 2001). The idea, as I understand it, is that physical laws, such as laws of quantum physics, can be applied to most objects regardless of their spatiotemporal locations, while a biological law can only apply to a limited number of composite objects. Another common explanation is that higher-level laws in special sciences can be multiply realised by different lower-level natural kinds (see, for example, Fodor 1974). For example, economic laws can be realised by computer data as well as paper notes.

Certainly, McDowell considers the space of reasons and our second nature to be non-natural phenomena. This might differentiate his view from those of the philosophers mentioned above but not from other supervenient metaphysical naturalists and physicalists, such as Ralph Wedgwood, who considers normative facts to be non-natural (2007, pp. 197-198). The views of the metaphysical naturalists and physicalists mentioned above are all controversial, but what is important here is that it is unclear how McDowell’s view of the irreducibility of the space of reasons and our second nature is different from those views that are widely held by metaphysical naturalists and physicalists. In sum, it is hard to see why McDowell’s ‘liberal naturalism’ should be considered a member of what De Caro and Voltolini consider to be ‘liberal naturalism’ but not fall under the umbrella of standard metaphysical naturalism.

A.9. Reasons for understanding the doctrines as separate

I have introduced the six doctrines that bear the name ‘naturalism’ and their variants. Some philosophers discuss several of the doctrines together as if they are components of one single grand doctrine, ‘naturalism’ (see, for example, Sterelny 1990, p. ix; Harman 2012, p. 10). For example, Gilbert Harman (2012) sees methodological and metaphysical naturalism as two sides of the same coin. He even defines metaphysical naturalism by methodological naturalism:

Naturalism also often has an ontological or metaphysical aspect in supposing that the world is the natural world, the world that is studied by the natural sciences, the world that is available to methodological naturalism. But the main naturalistic theme is methodological. (2012, p. 10)
On the other hand, some philosophers tend to think the doctrines are separate (see, for example, Devitt 1998, p. 46; Horst 2009, pp. 221-222; De Caro & Macarthur 2010, p. 4; Papineau 2014, p. 116; 2015; Clark 2015, pp. 1-5). For example, David Papineau understands metaphysical and methodological naturalism to be ‘largely independent’ doctrines:

Many different ideas parade under the banner of philosophical naturalism. […] ‘[M]ethodological naturalism’ is to be distinguished from ‘ontological naturalism’. […] While both ontological and methodological naturalism claim a species of affinity between philosophy and science, the two doctrines are largely independent. (2014, p. 116)

Michael Devitt even considers the two doctrines to be ‘sharply different’ (1998, p. 46).

I side with the latter group of philosophers. As we have seen, even though the doctrines sometimes motivate (or entail) one another, they have tremendously different meanings and often have their places in different philosophical areas. In addition, they are often incompatible with each other. Here are a few examples. The liberal naturalist establishes herself as a critic of methodological naturalism and the standard version of metaphysical naturalism. The indispensability naturalist’s use of our intuition is incompatible with the methodological naturalist’s exclusive methodological commitment to the scientific method. The claim of some indispensability naturalists that mental properties are immaterial is incompatible with metaphysical naturalism. In sum, I suggest understanding the doctrines as separate, despite the fact that they bear the same name.
Appendix B. The historical path approach to defining metaphysical naturalism and its rivals

In Section 2, I proposed and defended the historical path approach (HPA) to the disciplinary characterisation of metaphysical naturalism, but there are other attractive approaches to the disciplinary characterisation in the literature. They, much like HPA, have the potential to resist the objections to the disciplinary characterisation that I discussed in Section 2. In the main body of this dissertation, for the sake of having a more focused discussion, I set those rival approaches aside because they are not directly irrelevant to our purpose. Nevertheless, they are still worth of being discussed separately in this appendix. In this appendix, I will very briefly sketch three such approaches, including (1) the extra condition approach, (2) the purpose approach, and (3) the structurisation approach, and will argue that they are either less attractive than HPA or not really competitors of it.

Despite the fact that I do not believe that one single true definition of metaphysical naturalism has to be singled out, I still consider it to be useful to have a comparison between HPA and other approaches here. This is because for HPA to be preferable, it should still at least be as attractive as other approaches but not less attractive than them.

One last point before we begin. As I have noted in Section 2, metaphysical naturalism and physicalism are closely related, and hence arguments concerning the characterisation of one doctrine often apply to both doctrines. Much like the discussion in Section 2, I will focus on metaphysical naturalism, but I will nevertheless take arguments concerning physicalism into account.

B.1. The extra condition approach

The extra condition approach adds extra conditions other than the disciplinary characterisation into the definition of metaphysical naturalism to formulate cluster definitions. Jessica Wilson (2006), for example, suggests that the extra condition that there is no fundamental mentality can be added to the definition of physicalism. This approach falls back into the cluster approach I criticised in Section 2.3, and they share the same weaknesses. To repeat my criticism, the problem facing cluster definitions is that they are often arbitrary and ad hoc. It seems that their various conditions are independent, and it is difficult to see why they should be considered as a single doctrine. What motivates their defenders to add additional conditions into the single doctrine seems to be nothing more than their intention to exclude whatever they do not approve of. Many, myself included, find
this subjectivity and arbitrariness unsatisfactory because MN (and/or physicalism) is expected to be a non-arbitrary, systematic, and substantive metaphysical doctrine or framework (see, for example, Stoljar 2010, p. 88).

**B.2. The purpose approach**

The purpose approach suggests that physics has its purpose, and that only a particular kind of entity can thereby be seen as its target of investigation. Whether physicalism is true is determined by whether every ultimate supervenience base is a member the kind. While it is possible that future physics finds this purpose to be mistaken, physicalism in that case is proven false. For example, Philip Pettit (1993) considers physics to be the study of microphysical entities; and J. L. Dowell (2006) considers physics to be the study of ‘relatively fundamental elements’ (p. 39). Both ideas suggest that small-sized atomistic space-time occupiers constitute everything (Pettit 1993, pp. 213-214; Dowell 2006, p. 39).

Metaphysical naturalism and physicalism have to be separately assessed here, because natural science in general does not share all purposes or constraints of physics. It can be questioned whether the purpose approach can be applied to metaphysical naturalism. The answer is perhaps yes, even though I consider the resulted version of metaphysical naturalism to be an indefensible one. There are few, such as Arthur Danto (1967) and Steve Clarke (2009), who suggest that the scientific method could be seen as the purpose of science, and that metaphysically natural entities should therefore be defined as whatever susceptible to investigations through the scientific method. But this obviously counts too much as metaphysically natural. It is difficult to see why the scientific method cannot be used to study souls, the fundamentally mental, God, angels, Aristotelian *teleos*, and so on. In fact, Arthur Danto (1967, p. 448) and Steve Clarke (2009, p. 132) do argue that those things should, indeed, be counted as metaphysically natural, but I suspect that metaphysical naturalism in this overly inclusive form would become a trivial and useless thesis.

Actually, even if the approach is applied only to physicalism, it will be problematic because it is in conflict with actual physical theories. As I have argued in Section 2.3, while microphysicalism might be an assumption of the old schools of physics, it is no longer the case. It is, of course, possible to look for a meaning of fundamentality or another purpose of physics that is alternative to Pettit and Dowell’s proposals. For example, one interesting view is that physics is the study of global laws, while special sciences are the study of non-global laws (see, for example, Loewer 2001). The idea, as I understand it, is that physical laws, such as laws of quantum physics, can be applied to most objects
regardless of their spatiotemporal locations, while a biological law can only apply to a limited number of composite objects. But it is unclear that any alternative (or their combination) can act as a good characterisation of physicalism. The global laws view, for example, cannot even exclude Berkeleyan Idealism because being within God’s mind is considered global. Another possible solution for the friends of the purpose approach is to bite the bullet and suggest that current physics has, indeed, violated physicalism. But then it is unclear in what respect the stated purposes can still be reasonably claimed to be the purposes of physics.

**B.3. The structurisation approach**

The structurisation approach increases the complexity of the conceptual structure of the disciplinary characterisation in order to resist objections. Andrew Melnyk (1997, 2003), for example, argues that physicalism should be seen as our current best scientific hypothesis. Whenever there is a new scientific theory, it gives birth to a new physicalism that is not identical to the original physicalism, and that new physicalism always acts as a new rival hypothesis to the original hypothesis. Despite the fact that he believes there are many physicalisms that are not identical to each other throughout history, he believes that there is an ancestry of physicalism. ‘Sociologically and historically speaking’, there is a ‘commonality’, a ‘spirit of physicalism’, ‘something that transcends particular formations’ (2003, pp. 14-15). In a later article (2009), he provides a characterisation of naturalism very different from his characterisation of physicalism. Naturalism is characterised as a philosophical paradigm shared by some philosophers, according to which there is no *sui generis*, in the sense of nomologically inexplicable agency. This seems to be explaining the ‘spirit of physicalism’ and the ‘commonality’ his earlier work is describing.

The structurisation approach can be said to be the family of approach that HPA belongs to, since HPA, like Melnyk’s hypothesis approach, avoids objections by increasing conceptual complexity. However, I consider the trade-off of Melnyk’s proposal to be much higher because of its loss of simplicity and its mismatch with our intuitions about and ordinary uses of the term ‘physicalism’. For example, we do not usually think there are many separate physicalisms, or consider physicalism and the ‘spirit of physicalism’ to be separate, and it is unclear what the ‘spirit of physicalism’ even is. In fact, Melnyk’s approach to characterising naturalism is very similar to the metaphysical paradigm thesis I outlined in Section 2.10.2. If HPA, which is simple and systematic and can match our ordinary use of the terms ‘physicalism’ and ‘naturalism’, can resist the same objections, then there is no need for other versions of the structurisation approach.