

Revelation and the Intuition of Dualism

Michelle Yuanbo Liu



Somerville College
DPhil in Philosophy

REVELATION AND THE INTUITION OF DUALISM

Michelle Yuanbo Liu (Somerville College)

A Thesis Submitted in Trinity Term 2019 for the Degree of DPhil in Philosophy

ABSTRACT

There is a broad consensus among philosophers that when we think about the nature of consciousness, there is a persistent and plausibly widespread intuition of dualism – a belief or a disposition to believe that consciousness is nonphysical. This dissertation addresses the intuition of dualism and what explains it. The main body of the dissertation focuses on expounding a rational explanation of our dualistic intuition by drawing on what is known as the ‘thesis of revelation’ in the philosophy of mind, the claim that the essences of phenomenal properties are revealed in phenomenal experiences. The thesis of revelation, as I clarify, is incompatible with physicalism. Furthermore, as I argue, it is part of our ordinary conception of experience and its intuitiveness is what underlies our intuition of dualism. In my dissertation, I also defend the thesis of revelation from existing and potential objections.

The dissertation is divided into three main parts. Part I introduces the theme of the dissertation (chapter 1), elucidates the intuition of dualism (chapter 2), and critically assesses a number of candidate explanations of the intuition (chapter 3). Part II is dedicated to sharpening our understanding of revelation (chapter 4), clarifying its alleged incompatibility with standard versions of physicalism (chapter 5), and putting forward a linguistic argument for its intuitiveness (chapter 6). Discussion of these issues helps to bring forth a rational explanation for the intuition of dualism in terms of the intuitiveness of revelation (chapter 7). The issues of whether similar intuitions of distinctness can arise in other areas and of what explains the intuitiveness of revelation are also addressed (chapter 7). Parts I and II complete the main narrative of the dissertation. Part III deals with some further issues regarding revelation. It addresses objections against the revelation argument against physicalism and the thesis of revelation itself (chapter 8). It also provides a preliminary defence of property dualism, a position compatible with revelation, from the exclusion problem (chapter 9).

ACKNOWLEDGEMENTS

In 2013 during my second year in the BPhil programme in Oxford, I studied Chalmers' hard problem of consciousness for the first time. The problem of consciousness has preoccupied my own consciousness ever since. I have always had an unshakable intuition of dualism, though I always felt uneasy about rejecting physicalism. This dissertation is in a way a self-diagnosis of my dualistic intuition. I think it is also a plausible diagnosis of the intuition of dualism that many fellow philosophers share.

I am indebted to a great number of people who have helped me in one way or another in completing my dissertation. First and foremost, I would like to thank Martin Davies and Gonzalo Rodriguez-Pereyra for being such great and supportive supervisors – I have learned a great deal from both. Without them, this dissertation wouldn't have been possible. I am also extremely grateful to my college advisor Karen Margrethe Nielsen who has given me valuable support throughout my time at Somerville College.

I also benefitted greatly from various conferences and summer schools I attended over the years, especially from helpful conversations I had on these occasions. I had extremely helpful discussions with Philip Goff on several occasions, including at the *Matter, Mind and Consciousness Summer School* in Budapest, the *Scientism and Consciousness Conference* at Keele and most recently at the *Oxford Literary Festival*. I would like to thank my commentators Bill Child and Laura Gow for valuable comments on parts of the thesis which I presented at the *DPhil seminar* in Oxford and *MindGrad* at Warwick respectively. In 2017, I had a chance to attend the wonderful *Colour Primitivism Workshop* in Canada, organised by Derek Brown, from whom I also received helpful written feedback. I was also very lucky to be invited by Martine Nida-Rümelin to give a talk at the *Phenomenal Consciousness and Self-Awareness* workshop in Fribourg, Switzerland, at which I had fruitful discussions and very useful feedback.

In 2018, I took a two-term leave to spend a semester at the Australian National University. I am extremely grateful to my host, Daniel Stoljar, who helped me a great deal to shape the dissertation and hone some of the key arguments in it. I would also like to thank all the wonderful friends I made at the ANU, who made my time there a precious memory.

Finally, I give my eternal gratitude to my partner Luke King-Salter, my mother Zou Qing, my father Liu Xiangji, and my supervisor Martin Davies. I am truly indebted to these people. Martin has been the best supervisor one could ask for during both my BPhil and DPhil, and in both Oxford and Canberra. My family, whose kindness I will never be able to repay, have been most supportive and understanding.

CONTENTS

CHAPTER 1. INTRODUCTION	1
1. Conscious Experience	1
2. Outline	10
PART I	
CHAPTER 2. THE INTUITION OF DUALISM	15
1. Introduction	15
2. Physicalism and the Problem of the Explanatory Gap	16
2.1. Physicalism	16
2.2. Bridging the Explanatory Gap	23
2.3. A Disanalogy	31
2.4. From the Problem of the Explanatory Gap to the Intuition of Dualism	34
3. The Intuition of Dualism	39
3.1. Defining the Intuition of Dualism	39
3.2. Varieties of Intuitions of Dualism	42
4. Chapter Conclusion	46
CHAPTER 3. EXPLAINING THE INTUITION OF DUALISM: SOME PROPOSALS	47
1. Introduction	47
2. Adequate Explanation	47
3. Dual-Process Cognition	54
4. The Irreducibility of Subjectivity	59
5. The Antipathetic Fallacy	62
6. Conclusion of Part I	69
PART II	
CHAPTER 4. REVELATION	71
1. Introduction	71
2. On Essence	73
3. Lewis on Revelation	77
4. The Formulation	80
5. Knowledge <i>De Dicto</i> versus Knowledge <i>De Re</i>	88
6. Revelation and the Simplicity of Qualia	92
7. Chapter Conclusion	95
CHAPTER 5. REVELATION AND PHYSICALISM	96
1. Introduction	96
2. Lewis on Revelation and Physicalism	98
3. Revelation and Functionalism	100
4. The Revelation Argument against Physicalism	108
5. Chapter Conclusion	113

CHAPTER 6. THE INTUITIVENESS OF REVELATION: A LINGUISTIC ARGUMENT	115
1. Introduction	115
2. The Argument	120
3. Premise I	124
3.1. X Is (Essentially/Really/Actually/In Fact) Y	125
3.2. Know What X Really Is	127
3.3. Explaining the Data	131
4. Premise II	138
5. Objections	140
6. Chapter Conclusion	144
CHAPTER 7. EXPLAINING THE INTUITION OF DUALISM: REVELATION	146
1. Introduction	146
2. The Explanation	147
3. Objections	155
4. Revelation about Colour	159
5. Why is Revelation Intuitive?	168
6. The Qualitative Inaccuracy Hypothesis	175
6.1. QIH and the Intuitiveness of Revelation	179
6.2. Motivating QIH	180
6.2.1. The Analogy with Colour	181
6.2.2. The Anaesthetic Case	182
6.3. Against QIH	184
6.3.1. First Horn: Introspective Representations as Experiences	185
6.3.2. Second Horn: Introspective Representations as Beliefs	191
7. Chapter Conclusion	196
PART III	
CHAPTER 8. REVELATION AND OBJECTIONS	198
1. Introduction	198
2. Partial Revelation	200
3. 'Proves Too Little'	203
4. 'Proves Too Much'	211
5. The Case of Moore	215
6. Introspective Humility	226
7. Chapter Conclusion	235
CHAPTER 9. REVELATION AND DUALISM	237
1. What We Have Achieved So Far	237
2. Dualism and the Exclusion Problem	246
2.1. Árnadóttir and Crane	252
2.2. List and Stoljar	257
3. Chapter Conclusion	263
BIBLIOGRAPHY	265

CHAPTER 1

INTRODUCTION

1. CONSCIOUS EXPERIENCE

Our conscious experience is something we are intimately familiar with. It makes up an important part of our mental life. In philosophical discussion, especially in the context of the physicalist-dualist debate about the nature of consciousness, there is a prevalent idea that the philosophical notion of conscious experience is a notion that everyone can easily grasp. Philosophers, nonetheless, often appeal to examples to clarify the notion. For instance, in *The Conscious Mind*, Chalmers writes (1996: 4):

Trying to define conscious experience in terms of more primitive notions is fruitless. ... The best we can do is to give illustrations and characterizations that lie at the same level. ... I presume that every reader has conscious experiences of his or her own. If all goes well, these characterizations will help establish that it is just *those* that we are talking about.

... Conscious experiences range from vivid color sensations to experiences of the faintest background aromas; from hard-edged pains to the elusive experience of thoughts on the tip of one's tongue; from mundane sounds and smells to the encompassing grandeur of musical experience; from the triviality of a nagging itch to the weight of a deep existential angst, from the specificity of the taste of peppermint to the generality of one's experience of selfhood. All these have a distinct *experienced quality*. All are prominent parts of the inner life of the mind. (second emphasis added)

Chalmers' point is that we all have conscious experiences, and we can easily grasp the idea of *conscious experience* by way of examples. We all know what it is like to see different colours, what a nagging itch feels like, what it is like to listen to a favourite

piece of music, etc. Pondering on these examples introspectively, we can notice something common to all these experiences: they have some *experienced qualities* attached to them. These experienced qualities are what philosophers call the 'phenomenal properties' of experience. Other synonymous terms include 'qualia', 'subjective characters', 'phenomenal characters', 'phenomenal qualities', 'raw feels', 'what-it-is-likeness', etc. To be conscious is just to have experiences with these phenomenal properties. To explain consciousness¹ is then to explain why our experiences have these phenomenal properties.

A conscious experience is a mental event that a subject undergoes. When philosophers of mind talk about conscious experience, they sometimes use the phrase 'phenomenal state'. The notion of *state* in this context can be understood in the same way as the notion of *event*.² Events occupy particular points in time and may be extended in time. When a subject undergoes a conscious event, say, an experience of pain, the subject feels a certain way. There is a way the subject feels in virtue of being in pain. That is,

¹ Throughout the dissertation, by 'consciousness', I only mean *phenomenal consciousness*, not *access consciousness* which is 'availability for use in reasoning and rationally guiding speech and action' (Block 1995: 227).

² As Kim notes: 'in discussing the mind-body problem, philosophers speak indifferently in terms of events, states, and processes' (1966: 232). It is also important to bear in mind that some philosophers do distinguish these terms. For instance, Lewis (1995: 140) takes conscious experiences to be *events* of experiencing, which are havings of phenomenal *states* (see Lewis 1995: 140). One might also think that 'event' suggests something happening or changing; whereas 'state' suggests simply being in a particular condition and does not suggest something happening or changing. Nevertheless, the terms can be used in ways that go against these initial suggestions, e.g. 'During the whole event, absolutely nothing happened'; 'The committee is in the state of turmoil'.

there is something that it is like for the subject to be in that conscious event or state.³ Let's call properties such as *feeling a certain way*, 'experiential properties'. Experiential properties are properties of experiencing *subjects*. We can also speak of 'phenomenal properties' of experiential *events*, i.e. the what-it-is-likeness of conscious experiences. A conscious experience has a phenomenal property Q just in case the subject of the experience instantiates an experiential property Q'. John's experience has the phenomenal property, *the painfulness of pain*, in virtue of the fact that John has the experiential property of *feeling thus-and-so*.⁴

What it is like for a subject when she experiences pain is different from what it is like for the subject when she experiences an itch, or feels hot, or hears middle C. Conscious experience-tokens belong to the same conscious experience-type or phenomenal type insofar as the experience-tokens have the same phenomenal property. For example, I am having the conscious experience of hearing middle C currently. A musician might have the same kind of conscious experience every day. These

³ To say that there is something it is like for S to Φ is to say that there is something it is like *to S for S to Φ* , where S is both the subject of the psychological state being described by the phrase as well as the subject of Φ -ing (see Stoljar 2016).

⁴ Philosophers of mind predominantly speak of *experiences* as having 'phenomenal properties'. Sometimes, they speak of *subjects* having 'phenomenal properties', e.g. Chalmers (2003: 250) writes, 'whenever a subject has a phenomenal property, the subject is acquainted with that phenomenal property'. Here I have made a terminological distinction between *phenomenal* properties, which are properties of conscious mental *events*, and *experiential* properties, which are properties of conscious *subjects*. See Horgan (1984) for a similar distinction. Nida-Rümelin (2018) has argued against the *event-property framework* according to which experiences, as events, instantiate phenomenal properties. In my thesis, I shall sidestep the controversy and adopt the orthodox event-property framework.

numerically distinct experience-*tokens* are of the same phenomenal *type* insofar as they have the same phenomenal property, i.e. what it is like for the subject to hear middle C.

Given the clarification above, the term ‘conscious experience’ or ‘phenomenal state’ can mean either a *conscious experience-token* or a *conscious experience-type*. Since the notion of *type* can be understood as the notion of *property* or *universal*, the notion of *conscious experience-type* or *phenomenal type* can be understood as equivalent to the notion of *phenomenal property*. Indeed, philosophers of mind often use terms like ‘phenomenal state’, ‘conscious experience’, or simply ‘consciousness’ or ‘experience’ to mean *phenomenal property* (see Sundström 2011: 268). This meaning of ‘experience’ or ‘conscious experience’, i.e. *conscious experience-type* or *phenomenal property*, is the most important to the discussion on the metaphysics of consciousness and is the one we will be focusing on in this thesis.⁵

Physicalists tell us that our conscious experiences are just physical states, e.g. brain states.⁶ Brain state-tokens belong to the same brain state-type insofar as the state-

⁵ Linguist Anna Wierzbicka (2010) notes that the English word ‘experience’ has undergone significant semantic changes. In Shakespearean times, it predominantly meant ‘knowledge by living and doing things’ and this meaning is still present in contemporary usage of the word, e.g. teaching *experience*, an *experienced* driver, etc. However, in the 18th century, possibly due to the influence of British empiricism, the word took on a new reference to subjective aspects of mental episodes (Wierzbicka 2010: 39). So examples frequently used by philosophers such as ‘seeing red’, ‘hearing sounds’, ‘feeling anxious’, etc. are experiences in this post-17th-century meaning of the word. Wierzbicka (2010: 39) also notes that there is no single word in other European languages whose meaning perfectly matches the post-17th-century meaning of the English word ‘experience’. German, for instance, has two words which each cover part of the semantic territory of the English word ‘experience’: *Erfahrung* and *Erlebnis*. Emotions like sadness and joy are *Erlebnisse*, whereas sensations (or ‘impressions’) like seeing red, smelling flowers, etc. are *Erfahrungen*. This was also noted by Elizabeth Anscombe in translating Wittgenstein’s *Remarks on the Philosophy of Psychology* (1980, vol. 1: 839).

⁶ Here ‘physical states’ is understood in a broad sense to include functional states. For instance, according to standard versions of physicalism, the experience of pain is either the physical state

tokens have the same physical property. When physicalists claim that conscious experiences are physical states, the important sense of the physicalist identity statement is that conscious experience-types are identical to physical state-types, i.e. phenomenal properties are identical to physical properties (including functional properties).⁷

However, intuitively, conscious experiences don't seem to be what physicalists take them to be, that is, phenomenal properties don't seem to be physical properties. It seems surprising that my being in the phenomenal state of hearing middle C is just my being in some brain state. It seems false that the what-it-is-likeness of hearing middle C is just the physical property of being some physical event in the brain. It seems more intuitive to say that the kind of conscious experience I have when I hear middle C is distinct from but co-occurs with the kind of physical event that takes place in my brain whenever I hear middle C. It seems more intuitive to say that phenomenal properties are *not* physical properties.

The belief or the disposition to believe that conscious experience or consciousness is nonphysical is what has been called 'the intuition of dualism' (e.g. Papineau 2008, 2011, forthcoming; Chalmers 2018). Physicalists themselves have widely acknowledged the existence of such an intuition (e.g. Lewis 1995; Perry 2001; Papineau 2002, 2008, 2011, forthcoming; Braddon-Mitchell 2007; Sundström 2017). As John Perry has dramatically

of C-fibres firing or a functional state that is multiply realised by different physical states such as C-fibres firing. More on the notion of 'the physical' in chapter 2.

⁷ On Davidson's Anomalous Monism (1969, 1970), token-identity of the mental and the physical does not entail a corresponding type-identity. Here we are assuming that phenomenal properties are genuine properties of experience-tokens and individuate experiences into phenomenal types.

remarked: 'To say that this, the feeling I am aware of when I, so to speak, look inward, is that, the thing [some brain state] I read about, just seems crazy' (2001: 4). This intuition makes physicalism seem intuitively false, and its main alternative property dualism seem attractive.⁸ Indeed, there is a widespread consensus among both physicalists and dualists that while physicalism seems to be a satisfactory doctrine with respect to non-conscious phenomena, the existence of conscious experience seems to be uniquely hard to reconcile with the physicalist worldview; property dualism – which acknowledges the uniqueness of conscious experience – is, in a sense, the more intuitive position to take with respect to conscious experience.

The intuition of dualism has also featured in a recent paper by Chalmers (2018: 6) which discusses the meta-problem of consciousness, 'the problem of explaining why we think that there is a problem of consciousness'. The meta-problem of consciousness is important in the sense that 'a solution to the meta-problem will shed significant light on the hard problem [of consciousness]' (Chalmers 2018: 8). It is possible that such a solution could dissolve the hard problem, in which case we *think* that there is a problem about how and why 'physical processes in the brain give rise to conscious experience' (2018: 6; see also Chalmers 1996), but there really isn't one. According to Chalmers, to solve the meta-problem of consciousness we need to explain why we have certain 'problem intuitions' with respect to consciousness. Chalmers (2018: 11) divides these problem intuitions into four main categories as below:

⁸ Property dualism is distinguished from substance dualism. In this dissertation, by 'dualism', I am mainly concerned with property dualism.

Perhaps the core intuitions for the meta-problem as defined are *explanatory intuitions* holding that consciousness is hard to explain. These include gap intuitions holding that there is an explanatory gap between physical processes and consciousness, and anti-functionalist intuitions holding that explaining behavioral functions does not suffice to explain consciousness. Closely related are *metaphysical intuitions*, including dualist intuitions holding that consciousness is nonphysical, and fundamentality intuitions holding that consciousness is somehow fundamental or simple. There are also *knowledge intuitions*: these include both first-person knowledge intuitions holding that consciousness provides special knowledge from the first-person perspective (like Mary's knowledge of what it is like to see red on leaving the black and white room), and third-person ignorance intuitions, such as the intuition that it is hard to know the consciousness of other people or other organisms (such as what it is like to be a bat). There are *modal intuitions* about what is possible or conceivable, including the "zombie" intuition that a physical or functional duplicate of us might lack consciousness and "inversion" intuitions, such as that someone else might be experiencing red when I experience green.

Chalmers takes 'explanatory intuitions' and 'metaphysical intuitions' to be the central cases of 'problem intuitions' that need to be explained in solving the meta-problem of consciousness. The intuition of dualism, the intuition that consciousness is nonphysical, belongs to the category of metaphysical intuitions. One might reasonably assume that while the problem intuitions may be closely related and there might be a common factor that underpins these intuitions, detailed explanations of different problem intuitions would differ depending on the precise content of the intuition at issue and the precise style of explanation employed.

This dissertation focuses on the *intuition of dualism*, the intuition that consciousness is nonphysical, and focuses on expounding a rational explanation for the intuition. For Chalmers (2018: 11), to explain the meta-problem of consciousness is to

explain certain verbal reports, or judgements, or dispositions to make certain reports and judgements about consciousness, and the kind of judgements Chalmers focuses on are 'judgments that arise prior to philosophical argument'. In the case of dualistic intuitions, examples of such judgements are:

- It is hard to see how consciousness could be physical.
- Phenomenal states are not brain states.
- It is intuitively false that phenomenal properties are physical/functional properties.
- The experience of pain doesn't seem to be the firing of C-fibres.⁹

As we have already clarified, terms like 'consciousness', 'phenomenal states', 'phenomenal properties', 'experience', etc. are used interchangeably in the literature and in this context are taken to mean the same thing. Put simply, we can say that the intuition of dualism is the intuition that consciousness, that is the phenomenal properties of experience, are nonphysical.

Genuine utterances like the above can be seen as expressions of beliefs. Our attributions of beliefs are governed by constraints of rationality (see Davidson 1984; Davies and Coltheart 2000). We treat people as rational agents and their beliefs and actions as governed by reasons. Similarly, we expect our intuition of dualism, our belief

⁹ Note that these examples, except the first on the list, are not from Chalmers (2018). But one can find similar expressions of the intuition of dualism in Perry 2001; Papineau 2011; Robinson 2015; Sundström 2017.

or disposition to believe that consciousness is nonphysical, to exhibit a certain kind of reasonableness.

This dissertation focuses on expounding a rational explanation of our intuition of dualism. In putting forward this explanation, I draw on what is sometimes known as ‘revelation’ in the literature on philosophy of mind (e.g. Stoljar 2009; Chalmers 2016, 2018; Goff 2017). The thesis of revelation is the claim that the essences of phenomenal properties are revealed in phenomenal experiences. There has been much discussion on revelation in recent literature on the metaphysics of consciousness (Goff 2015, 2017; Trogon 2017; Majeed 2017; Stoljar 2018). Notably, it has been used to argue against physicalism (Goff 2017; see also Nida-Rümelin 2007; Horgan and Tienson 2001). Nevertheless, what the thesis precisely states and what role it is supposed to play in the debate about the metaphysics of consciousness calls for clarification and in-depth discussion. Much of this dissertation is dedicated to addressing these issues pertaining to revelation. So, in addition to explaining our intuition of dualism in terms of the intuitiveness of revelation, the dissertation also makes an attempt at spelling out the important role that the thesis of revelation plays in the metaphysical debate about the nature of consciousness. Overall, I contend that the thesis of revelation is intuitive in the sense of being part of our ordinary conception of experience, and that the intuitiveness of revelation is the rationale that underpins our intuition of dualism. Furthermore, I argue that the thesis is also a defensible and tenable philosophical position. The following section provides an outline of the dissertation and a brief summary of each chapter.

2. OUTLINE

The dissertation is divided into three parts. Part I elucidates and clarifies the intuition of dualism, and critically assesses a number of candidate explanations of the intuition. Part II, which forms the most important part of the dissertation, offers a careful articulation of the thesis of revelation and puts forward an explanation of the intuition of dualism in terms of the intuitiveness of revelation. Part III focuses on metaphysical questions one might raise regarding the thesis of revelation – it defends revelation from objections and discusses the prospects of dualism, which is compatible with revelation. A more detailed plan of each chapter is as follows.

Chapter 2 ‘The Intuition of Dualism’ begins with a discussion of physicalists’ attempts to solve the problem of the explanatory gap, the problem of explaining conscious experience in physical terms. This discussion brings out a disanalogy between phenomenal-physical identity claims such as ‘the experience of pain is the firing of C-fibres’ and other scientific identity claims such as ‘water is H₂O’. Whereas we are not puzzled about the latter, we persist in finding the former incredible. This disanalogy makes evident the existence of an intuition of dualism. This intuition is defined as the belief or the disposition to believe that consciousness is nonphysical. Such an intuition is persistent and widely shared. In clarifying what I take to be the intuition of dualism at issue, I also clarify how the intended notion of the dualist intuition is to be distinguished from the kinds of ‘dualistic intuitions’ that permeate discussions in anthropology and developmental psychology.

Given that there is an intuition of dualism in the intended sense, what then explains it? Chapter 3 ‘Explaining the Intuition of Dualism: Some Proposals’ surveys and assesses a number of existing proposals for explaining the intuition of dualism: (a) *dual-process cognition* – an account from Fiala, Arico and Nichols (2011) which explains the intuition of dualism by appealing to dual-process cognitive architecture; (b) *the irreducibility of the subjective* – a proposal suggested by Nagel (1986) according to which the subjective and the objective do not seem to converge on a common reality; and (c) *the antipathetic fallacy* – an account put forward by Papineau (2002) which explains the intuition of dualism in terms of the uniqueness of the phenomenal concepts that we employ in thinking about conscious experiences. I argue that these proposals for explaining the intuition of dualism all fail to be satisfactory.

The failure of these proposals leads me to look for an alternative explanation. Part II of the dissertation focuses on the thesis of revelation, which claims that the essences of phenomenal properties are revealed in experience. Recent discussion on the metaphysics of consciousness has seen a surge in the discussion on revelation (see Goff 2017; Majeed 2017; Trogon 2017; Stoljar 2018). While mentions of revelation are frequent, there is room for further discussion of how precisely to formulate the thesis of revelation and what exactly it amounts to. While I aim to use ideas discussed in Part II to bring out a plausible explanation of the intuition of dualism in terms of the intuitiveness of revelation, I also intend this part of the dissertation to contribute to the existing literature on the thesis of revelation and its role in the debate on the metaphysics of consciousness.

Chapter 4 ‘Revelation’ introduces and clarifies the thesis of revelation. A precise formulation of the thesis is as below:

(Revelation)

By having an experience-token with phenomenal property Q,
one is in a position to know a truth, namely, ‘Q is X’, where
the predicate ‘X’ captures the essence of Q, although it may be
hard to put into words.

A detailed explanation of this formulation will be provided in chapter 4.

In the literature, the thesis of revelation is often said to be incompatible with physicalism (see Lewis 1995; Goff 2017), but not everyone agrees on this point (see Damjanovic 2012; Trogon 2017; Stoljar 2018). Chapter 5 ‘Revelation and Physicalism’ clarifies how the incompatibility between the thesis of revelation and physicalism is precisely understood. Physicalists standardly claim that phenomenal properties have physical or functional essences. Given revelation, physicalism entails that in having an experience with phenomenal property Q, one is in a position to know that ‘Q is X’, where ‘X’ is a physical or functional predicate which captures the essence of Q. I shall argue that in having an experience with Q, one is not in a position to know such a truth, and that the thesis of revelation is incompatible with what I call ‘standard physicalism’, which claims that qualia have physical or functional essences.

Chapter 6 ‘The Intuitiveness of Revelation: A Linguistic Argument’ turns to the intuitiveness of revelation. Many, including physicalists (Lewis 1995: 142; McLaughlin 2003: 99; Braddon-Mitchell 2007: 287; Hill 2014: 199-200; Papineau forthcoming), take the

thesis of revelation to have an intuitive appeal. Notably, revelation is often thought to be intuitive in the sense of being part of the ordinary conception of experience (Lewis 1995; Braddon-Mitchell 2007). But the claim that revelation is part of our ordinary conception of experience is controversial (Stoljar 2009). Chapter 6 advances a linguistic argument, contending that the hypothesis that revelation is part of the ordinary conception of experience best explains some of our linguistic intuitions concerning experience. By inference to the best explanation, revelation is part of the ordinary conception of experience and is, in that sense, intuitive.

Building on chapters 4, 5 and 6, chapter 7 'Explaining the Intuition of Dualism: Revelation' turns to the question left unanswered at the end of Part I – What explains the intuition of dualism? This chapter puts forward an explanation for the intuition of dualism that appeals to the intuitiveness of the thesis of revelation. According to the proposed explanation, in judging that phenomenal properties are not or do not seem to be physical properties, we draw on our tacit or implicit understanding of the thesis of revelation, which is part of our ordinary conception of experience, as well as a tacit or implicit appreciation of the entailment from the thesis of revelation to the claim that phenomenal properties are not physical properties (and that a particular phenomenal property is not a physical property). On this proposal, the intuition of dualism has an inferential or inference-like aetiology, to which the subject need not have conscious access.

In addition, chapter 7 addresses two further issues. It discusses the thesis of revelation regarding colour and the issue of whether a similar intuition to the intuition

of dualism arises in the case of colour properties. It also delves into the question of what explains the intuitiveness of revelation. In particular, it looks into Derk Pereboom's (2011) Qualitative Inaccuracy Hypothesis as a physicalist proposal for explaining away the intuitiveness of revelation. I shall argue that this proposal fails.

Part I and Part II form the main narrative of the dissertation. Part III, the final part of the dissertation, addresses further questions one might raise regarding the thesis of revelation in the context of the metaphysics of consciousness. Chapter 8 'Revelation and Objections' is concerned with the force of the revelation argument against physicalism. In particular, it looks into Stoljar's (2018) review of Goff's book *Consciousness and Fundamental Reality*. The latter contains a version of the revelation argument against physicalism to which Stoljar raises three objections, including a direct objection against the thesis of revelation itself. I shall reply to Stoljar's objections. In addition, chapter 8 also relates revelation to the topic of introspection and clarifies how the limits of introspection raise no problem for the thesis of revelation.

Chapter 9 'Revelation and Dualism' concludes the dissertation and addresses the prospects of property dualism. While property dualism is compatible with revelation, it faces the causal exclusion problem. Chapter 9 provides a preliminary defence of property dualism against the causal exclusion problem. I argue that the problem does not put as much pressure on dualism as it is commonly thought, because it is unclear that the exclusion principle, once clarified, is well-motivated.

Chapter 2

THE INTUITION OF DUALISM

1. INTRODUCTION

The problem of the explanatory gap is usually understood as the problem that we lack a satisfactory explanation of conscious experience in physical terms. Advocates of the problem claim that we lack an answer to, for instance, the question of what it is about the firing of C-fibres that gives rise to the phenomenal experience of pain (Kim 2005: 93). The explanatory gap is commonly regarded as a stumbling block for physicalism and most physicalists deny the existence of the explanatory gap thus understood. A priori physicalists, for example, argue that we do indeed have an explanation of conscious experience in physical terms, and that this explanation involves giving conscious experiences functional definitions. Many physicalists, however, appeal to a posteriori identity claims in bridging the explanatory gap. These physicalists claim that phenomenal states are just physical states, and we have sufficient evidence to think so.

Despite various attempts made by physicalists to address the problem of the explanatory gap, the intuition that there is an explanatory gap lingers. In particular, while we don't question scientific identity claims such as 'water is H₂O', we seem to be resistant to phenomenal-physical identity claims such as 'the experience of pain in humans is the firing of C-fibres in the brain'. This disanalogy between phenomenal-physical identity claims and other identity claims suggests that there exists a deep intuition of dualism, the intuition that conscious experience is nonphysical.

This chapter motivates the claim that there exists an intuition of dualism and clarifies what precisely this intuition is. The structure of the chapter is as follows. Section 2 discusses the problem of the explanatory gap to make it evident that there exists a persistent intuition of dualism. Section 3 looks closer at the intuition of dualism. It clarifies what the intuition is and distinguishes it from the kinds of ‘intuitions of dualism’ discussed in psychology and social-cultural studies. Section 4 concludes the chapter.

2. PHYSICALISM AND THE PROBLEM OF THE EXPLANATORY GAP

In this section, I clarify the doctrine of physicalism and address the problem of the explanatory gap, as well as various physicalists’ attempts to solve the problem. Discussion in this section demonstrates that there exists an intuition of dualism when we think about consciousness.

2.1. PHYSICALISM

While talk of physicalism is commonplace, a precise formulation of the doctrine has proven to be difficult (see Stoljar 2010). What I say here regarding physicalism will be relatively uncontroversial and will be sufficient to facilitate our discussion.

It is obvious that mental phenomena are intimately associated with physical phenomena – the mental correlates with and depends on the physical. Physicalists take

mental properties to *metaphysically supervene* on physical properties.¹ Here the supervenience base includes all physical properties, relations and fundamental physical laws. The supervenience base necessitates the supervenient, where the force of necessitation is understood to be metaphysical – necessity in the strictest sense – as opposed to nomological.² So physicalism entails that physical properties *metaphysically necessitate* or are *metaphysically sufficient* for chemical, biological, economic and all other kinds of properties including phenomenal ones. To put it another way, we can say that physicalism entails that the conditional ‘If P then C’ (where P stands for the totality of physical facts and C a given phenomenal fact) is necessarily true.

But the notion of *metaphysical supervenience* is insufficient for capturing the doctrine of physicalism. This inadequacy is made salient by dualistic metaphysical theories like necessitarian dual attribute theory (see Jackson 2006). According to these theories, the phenomenal is fundamentally distinct from but metaphysically supervenes on the physical. Such supervenience is brute and not explainable solely in terms of the physical. It would seem that in order to distinguish physicalism properly understood

¹ The notion of *supervenience* commonly used in formulating physicalism is that of *global supervenience* – all properties globally supervene on physical properties if and only if any possible world *w* which is a minimal physical duplicate of our world is a duplicate of our world *simpliciter* (Jackson 1998: 12; Chalmers 1996: 42). Physicalism is compatible with the possibility of ‘epiphenomenal ectoplasm’ – a pure psychological phenomenon that does not causally interact with the rest of the world – but not with its actuality. So it is inadequate to formulate physicalism as the claim that ‘any world that is a physical duplicate of our world is a duplicate of our world *simpliciter*’. A *minimal* physical duplicate of our world is meant to be one which is exactly like our world in every physical respect and contains nothing else, i.e. does not include the existence of epiphenomenal ectoplasm.

² One can spell out the difference in terms of possible worlds. Metaphysical necessities are truths which hold in all possible worlds; nomological necessities are truths which hold in all worlds with the same laws as the actual ones.

from necessitarian versions of dualism, physicalists need to explain why the physical by itself metaphysically necessitates the phenomenal (see Horgan 1993; Morris 2018). They would have to answer questions such as: What is it about brain states that accounts for their correlation with phenomenology? Why is having C-fibres firing metaphysically sufficient for experiencing pain?

To answer these questions is precisely to face up the problem of the explanatory gap, which is the challenge of explaining conscious experience in physical terms. The problem engenders two kinds of physicalism, *a priori* and *a posteriori*, which correspond to two kinds of strategies aimed at solving it. I will say more about these two kinds of physicalism in the next subsection when I explain the notion of reductive explanation and the problem of the explanatory gap.

Another notion that is crucial in defining physicalism is the notion of 'the physical'. The 'physical' commonly refers to properties and laws acknowledged by

physics.³ This is the narrow sense of the term.⁴ There is also a broad sense of the term. According to this broad sense, properties posited by, not only physics, but also the special sciences, e.g. chemistry, biology, neuroscience, etc., are considered ‘physical properties’. Indeed, physicalists usually identify phenomenal properties with properties posited by neuroscience, e.g. the experience of pain is C-fibres firing in the brain. Functional properties, that is causal-role or functional-role properties, also count as physical properties in the broad sense. Throughout the dissertation, I shall sometimes use the term ‘physical’ in the broad sense to include functional properties, and

³ Stoljar (2001) distinguishes between two senses of ‘the physical’ – a theory-based conception and an object-based conception. The theory-based conception is how we understand the notion of the physical here, where physical properties are defined with reference to physics. They are properties acknowledged by physics, that is, properties designated by physical predicates which feature in the vocabulary of physics. Call such a physical property ‘t-physical’. According to the object-based conception, physical properties are defined with reference to paradigmatic physical objects. Physical properties, under this conception, are properties required to give a complete account of the intrinsic nature of paradigmatic physical objects and their constituents or else properties which metaphysically supervene on the former sort of properties. Call such a physical property ‘o-physical’. All t-physical properties are also o-physical since the object-based conception, by definition, provides a complete account of the nature of physical objects. But it is not clear that all o-physical properties are necessarily t-physical (Stoljar 2001: 261). There could be properties which are not captured by the theory-based conception, but are nevertheless included in the object-based conception of physical properties. Here, we are adopting the theory-based conception of physical properties, which is the common conception of ‘the physical’ in the literature.

⁴ Since the notion of the physical goes hand in hand with physics, one might wonder what conception of physics we should adopt. If ‘physics’ is understood as current physics, then physicalism is likely to be false given that current physics is incomplete. If ‘physics’ is understood as complete physics, then everything would automatically count as ‘physical’, and physicalism would be trivially true. I think we need not worry about Hempel’s (1965) dilemma here. First, we could impose constraints on the notion of physics so that future physical theories will be sufficiently similar to current physics (Stoljar 2001: 257, n9). For instance, there should be constraints on the kind of fundamental properties that can be included in complete physics such that they should not include mental or proto-mental properties. Second, the problem posed by Hempel’s dilemma need not concern those physicalists who are confident that *current* physics already provides the resources needed to explain consciousness. Many physicalists would say that although we are still in search of neural correlates of various phenomenal states, it is not the case that the existing fundamental physical categories are inadequate to account for the physical nature of consciousness (Jackson 2004; Papineau 2002, 2011).

sometimes exclude them. This will either be made clear by the context or be specified explicitly.

It is also worth clarifying that in discussing physicalism, I will focus on what I call 'standard versions of physicalism' or 'standard physicalism'. According to standard versions of physicalism, which most physicalists adhere to, phenomenal states *are* physical states (including functional states), and the phenomenal properties of a token phenomenal state *are* purely physical properties – physical descriptions (including functional descriptions) exhaust the natures of phenomenal properties. In spelling out the incompatibility between revelation and physicalism in chapter 5, I will focus on the incompatibility between revelation and standard versions of physicalism. Recently, physicalism has also been defined in terms of the notion of *grounding* (e.g. Dasgupta 2014). Grounding is commonly understood as a non-causal explanatory relation that holds between facts. Physicalism, defined with reference to the notion of grounding, says that phenomenal facts are grounded in physical facts. Whether revelation is incompatible with grounding versions of physicalism is an issue I shall investigate in chapter 8.

The main argument for physicalism is the *causal argument*. It runs as follows (see Stoljar 2010; Papineau 2001, 2002):

- (1) All physical effects have physical properties as sufficient causes. (causal closure)
- (2) All mental properties have physical effects. (mental causation)

- (3) No two distinct properties can be sufficient causes for the same effect (except in cases of genuine overdetermination).
(the exclusion principle)
- (4) The physical effects of mental properties are not (systematically) overdetermined. (no overdetermination)
- (5) Mental properties are physical properties.

The causal argument for physicalism can be reformulated as *the exclusion argument* against dualism. We will discuss the latter in more depth in chapter 9, which should also shed light on the causal argument for physicalism. If the exclusion argument against dualism turns out to be problematic, then one might also question the force of the causal argument for physicalism.

For now, it is worth clarifying two points regarding the causal argument. First, the above version of the argument is formulated in terms of *properties*. There is also an event version of the argument according to which causal relata are thought of as events conceived as particulars (see Davidson 1969, 1993; Stoljar 2010: 211-2). The property version of the argument is more relevant than the event version for two reasons (see Stoljar 2010: 212). One, physicalism is understood as a thesis about properties, not particulars. Even a property dualist can agree that a particular mental event is identical to some physical event. So the event version of the causal argument cannot be an argument in favour of physicalism. Second, it is also intuitive to think that events cause other events in virtue of some properties being causally efficacious (see Kim 1993a). When I put my hand on the hot stove which causes me to experience pain, it is the *hotness*

of stove that is relevant to my feeling pain, not its size or colour. So it is intuitive to think that what is actually doing the causing are properties, not mere particulars.

It is also worth noting that the causal argument seems to be an argument only for identity physicalism (also known as 'reductive physicalism'), according to which mental properties are physical (not functional) properties; not for role functionalism, a commonly held version of nonreductive physicalism. On role functionalism, phenomenal properties are identical to functional properties, and the latter are distinct from but multiply realised by different physical properties. It seems that such functional properties are causally excluded by physical properties which play the relevant functional roles. Indeed, nonreductive versions of physicalism such as role functionalism also face the causal exclusion problem which is levelled against property dualism (see Kim 1993b; Papineau 2001). More on this in chapter 9.

Having clarified the doctrine of physicalism and the main argument for it, let us now turn to the problem of the explanatory gap. A closer look at physicalists' attempts to close the explanatory gap brings out the uniqueness of phenomenal-physical identity claims and our intuition of dualism.⁵

⁵ It is important to note that one can be committed to physicalism without taking the view that current science has bridged the explanatory gap. For instance, Stoljar (2006) has argued for the view that while it may be true that phenomenal truths follow a priori from physical truths, we are ignorant with respect to the kind of physical facts which would bridge the explanatory gap and make intelligible the conditional entailment 'If P then C'. In other words, one can be a physicalist and think that the explanatory gap will be bridged by future physics. However, most contemporary physicalists think that there is no explanatory gap in the sense defined at the beginning of this chapter. In the following, we will set Stoljar's view aside and focus on physicalists' proposals for bridging the explanatory gap. This move is unproblematic since the purpose of this section is not to argue that physicalists cannot bridge the explanatory gap or

2.2. BRIDGING THE EXPLANATORY GAP

The problem of the explanatory gap is the challenge of explaining how the physical by itself metaphysically, and not just nomologically, necessitates the phenomenal. To put it differently, it is the challenge of explaining conscious experience in purely physical terms. To bridge the explanatory gap, we need a reductive explanation of the occurrences of conscious states in terms of the occurrences of physical states. A reductive explanation involves two domains – a higher-level and a lower-level domain – and is an explanation of a higher-level phenomenon in terms of lower-level phenomena.⁶ The explanatory resources for a reductive explanation are drawn solely from the lower level, such that a reductive explanation of *A* may only refer to the lower-level phenomena *B* in its explanans (Kim 2005). So appealing to bridge laws, e.g. ‘whenever there are C-fibres firing in a subject, the subject experiences pain’, which do not only utilise explanatory resources from the lower-level phenomena, is unacceptable in a reductive explanation because the resultant explanation would not be reductive.

However, statements from two distinct levels normally look drastically different. How then do we facilitate the derivational transition from the lower-level explanans to the higher-level explanandum?⁷ There are two possibilities here (see Kim 2005: 107-8):

that physicalism is false, but to bring out the intuition of dualism, which will be made evident through a closer look at physicalists’ attempts to bridge the explanatory gap.

⁶ For instance, the reductive explanation of the boiling of a certain liquid is in terms of the lower-level phenomenon, the movements of molecules which make up the liquid.

⁷ Kim calls this ‘the problem of explanatory ascent’ (Kim 2005: 107). Here I follow Hempel among many others and assume that reductive explanations involve logical derivation.

- (a) *Conceptual Analysis*: providing a conceptual analysis of the higher-level explanandum in terms of the lower-level explanans.
- (b) *Identity Claims*: identifying the higher-level explanandum with a certain lower-level phenomenon.

Both (a) and (b) serve as definitions in addition to the empirical explanatory resources we have from the lower level. Definitions facilitate cross-level derivations. Option (a) provides a semantic analysis of the explanandum in a way that conceptually links concepts at different levels. Option (b) identifies some higher-level phenomenon with some lower-level phenomenon. The identity claim is supposedly established on independent grounds or based on theoretical considerations. Given these two methods of derivational transition, we have two corresponding models of reductive explanation of conscious experiences in terms of physical states. Option (a) is utilised by a priori physicalists and gives rise to the functionalist model of reductive explanation. Option (b) is utilised by a posteriori physicalists who appeal to relevant phenomenal-physical type-identity claims.

Let us first consider option (a) and the model of reductive explanation that appeals to a conceptual analysis of the explanandum. Consider the following reductive explanation for 'S has gene X at t':

- (I) Having gene X =_{def} having some biological mechanism with the causal role R, where R is the encoding and transmitting of genetic information Y.

- (II) DNA molecules *M*, found in human cells, have the causal role *R*.
- (III) DNA molecules *M* occur in *S*'s cells at *t*.
- (IV) *S* has gene *X* at *t*.

This reductive explanation starts with (I) where gene *X* is given a conceptual analysis in terms of its causal or functional role *R*. (II) involves a significant amount of scientific work at the level of molecular biology. The realisers of gene *X* at the basal level must fill the causal role *R*. Moreover, we should have a basal-level theory that explains how the realisers perform the causal role (Kim 2008: 132-3). We might come to identify DNA molecules *M*, which fulfil the causal role *C*.⁸ Upon observing the occurrence of *M* in *S* at time *t*, i.e. (III), we can predict and explain the occurrence of gene *X* in *S* at *t*, i.e. (IV). In this way, we can also make sense of the correlation between having gene *X* and having DNA molecules *M*. We can know what it is about *M* such that those molecules correlate with gene *X* – the former play the causal role which is constitutive of the latter.

Let us now consider how a priori physicalists would explain '*S* is having an experience of pain at *t*' using this model of reductive explanation:

- (i) Having an experience of pain =_{def} being in some physical state
with the causal role *C*.⁹

⁸ There could be alien life forms in which genetic information *Y* is encoded and transmitted in a different way, by molecules *N* instead of *M*. In such a case, the gene *X* is multiply realised.

⁹ On the functionalist model, there is a question about what the experience of pain is identical to. One might take it to be the physical state that realises the relevant functional role, e.g. experience of pain in humans = firing of C-fibres in the brain (e.g. Armstrong 1968; Lewis 1972). This is known as *realiser functionalism*. One might alternatively take the experience of pain to be identical to the functional state of being in *some* physical state or other with the causal role *C*,

- (ii) Firing of C-fibres plays the causal role *C* in humans.
- (iii) C-fibres are firing in human subject *S* at *t*.
- (iv) *S* is having an experience of pain at *t*.

The goal is to explain (iv). Step (i) of the model gives a functional definition for pain experience. The causal role *C* is supposed to be definitive or constitutive of pain experience. Step (ii) involves significant scientific work. The realisers of the experience of pain at the basal level must play the causal role *C*. Having identified the right physical realiser, in this case firing of C-fibres, we can predict and explain the occurrence of pain experience upon observing C-fibres firing in the same subject. So the explanandum (iv) follows from (ii) and (iii) via the relevant conceptual analysis (i). Because of the availability of conceptual analysis, any truths about phenomenal states follow from truths about their physical realisers. In this way, we can also make sense of what it is about the firing of C-fibres such that it metaphysically necessitates the conscious experience of pain – the former plays the causal role which is supposedly constitutive of the latter.

The problem with this model of reductive explanation is, right from the start, that we cannot give a conceptual analysis of phenomenal experience purely in functional terms without anything being left out (see Block and Fodor 1972; Block 1978). Intuitively, there is at least a conceptual difference between the causal role (described in physical terms) of a pain experience and the phenomenal character of a pain experience, i.e. the

which may be realised by different physical states in different physical systems, e.g. in differently constituted creatures (e.g. Putnam 1960, 1967). This is known as *role functionalism*.

painfulness of pain. Intuitively, being in a physical state with a particular causal role does not seem to guarantee having an experience with a particular phenomenal character (Levine 1983). Thus, a priori physicalists' strategy to bridge the explanatory gap by appealing to conceptual analysis and the functionalist model of reductive explanation seems highly implausible.¹⁰

Many physicalists, i.e. a posteriori physicalists, do not take reductive explanations to necessarily involve conceptual analyses. These physicalists opt for strategy (b). Recall that to facilitate the deductive transition from the lower-level explanans to the higher-level explanandum, this model of reductive explanation appeals to identity claims which identify the higher-level explanandum with a certain lower-level phenomenon. In the case of gene X, one might posit the following type-identity claim:

¹⁰ Functionalists like Jackson (2004) have appealed to representationalism to defend physicalism, and take the explanatory gap to be merely illusory. They argue from representationalism to physicalism by way of the following argument:

- (I) Phenomenal properties supervene on representational properties.
- (II) Representational properties supervene on physical properties.
- (III) Phenomenal properties supervene on physical properties.

Stoljar (2006: 212-3) has argued convincingly against this approach. I will briefly summarise his point here. Stoljar's argument depends on a distinction between *reductive* and *nonreductive* representational properties where the nature of the latter, but not the former, encompasses phenomenal properties (see also Chalmers 2010a, which distinguishes *pure* representational properties from *impure* representational properties. The former are properties of representing certain intentional contents. The latter are properties of representing certain intentional contents in certain ways.) It seems that it is only plausible that phenomenal properties supervene on the nonreductive kind of representational properties. But it is implausible that such representational properties supervene on physical properties. In other words, 'representational properties' in (I) should be understood as nonreductive representational properties, but this would render (II) false. (II) would be true only if one is already convinced of (III), in which case the above argument is circular.

(I) Gene *X* = molecules *M* in human cells.

According to advocates of this model of reductive explanation, identities like (I) are supposed to be established on independent grounds or based on theoretical considerations, rather than on some conceptual analysis of 'gene *X*'. In the case of the experience of pain, an a posteriori physicalist might posit the following phenomenal-physical type-identity claim:¹¹

(i) The experience of pain (in humans) = the firing of C-fibres in the brain

Consider then how appealing to (i) can give a reductive explanation for *S*'s having an experience of pain at *t*:

S is a human being and there are C-fibres firing in *S*'s brain at *t*.

(i) The experience of pain (in humans) = the firing of C-fibres in the brain

S is having an experience of pain at *t*.

If human pain can be identified with C-fibres firing in the brain, then given the occurrence of C-fibres firing in subject *S*'s brain, we can deduce and explain the occurrence of the experience of pain in the same subject. According to this model of reductive explanation, there is no problem of the explanatory gap – C-fibres firing alone is metaphysically sufficient for experiencing (human) pain because the two are one and the same thing.

¹¹ By 'phenomenal-physical identity', I include the identity of the phenomenal with the functional.

Now, regarding this strategy for bridging the explanatory gap, the crucial question is whether we have strong justification for positing relevant identity claims like (I) and (i). A posteriori physicalists argue that we have compelling reasons for positing them. They appeal to the following kinds of considerations (note that these considerations are by no means mutually exclusive):

(1) co-occurrence

(2) sameness in behaviour

(3) inference to the best explanation

Regarding (1) and (2), Papineau (2011: 10) writes:

Identities can be evidenced more directly in a number of ways. We might simply observe that the two kinds at issue *co-occur*. Or we might note that the physical kind in question *shares some behaviour* with the everyday kind. For example, it would be pretty good evidence that H₂O is water that it expands on freezing in just the way that H₂O does. (italics added, see also Papineau forthcoming; Smart 1959)

Regarding (3), Block and Stalnaker (1999: 23-4) write:

Why do we suppose that heat = molecular kinetic energy? Consider the explanation ... of why heating water makes it boil. Suppose that heat = molecular kinetic energy, pressure = molecular momentum transfer, and boiling = a certain kind of molecular motion. (...) Then we have an account of how heating water produces boiling. If we were to accept mere correlations instead of identities, we would only have an account of how something correlated with heating causes something correlated with boiling. Further, we may wish to know how it is that increasing the molecular kinetic energy of a packet of water causes boiling. Identities allow a transfer of explanatory and causal force not allowed by mere correlations. Assuming that heat = mke, that pressure = molecular momentum transfer, etc. allows us to explain facts that we could not otherwise explain.

Thus, we are justified by the principle of inference to the best explanation in inferring that these identities are true. (italics added)

In the case of ordinary phenomena such as heating and boiling, Block and Stalnaker think that we should opt for type-identities like heat = mke, etc., because positing the latter identity claims allows us to best explain various ordinary phenomena which we take to be genuine causal phenomena, e.g. heating water causes it to boil. They take such reasoning to apply also to the case of conscious experience. Within the mental domain, we also take various phenomena to be causal phenomena. We intuitively take a certain phenomenal type M_1 to cause a certain phenomenal type M_2 , e.g. frustration causes anger, pain experience causes distress, etc. We know that M_1 correlates with neurophysiological type N_1 and M_2 with N_2 . We also know how N_1 causes N_2 given our neurophysiological theory. If we posit $M_1 = N_1$ and $M_2 = N_2$, we can explain how M_1 causes M_2 . Block and Stalnaker would say that positing such phenomenal-physical type-identities *best explains* the fact that M_1 causes M_2 , e.g. pain experience causes distress. By inference to the best explanation, we are justified 'in inferring that these identities are true'.

A posteriori physicalists might also add that positing type-identities best explains the relevant co-occurrences and sameness in behaviour observed (Papineau forthcoming). For instance, they might say that the hypothesis that water is H_2O is the best explanation for the observed facts that water and H_2O co-occur and have the same behaviour. By inference to the best explanation, water is H_2O . According to these physicalists, something similar can also be said regarding phenomenal-physical identities.

Let's grant that positing type-identities allows the best explanations of ordinary causal phenomena in the physical domain, such as the fact that heating causes boiling, as well as relevant co-occurrences and sameness in behaviour. Let's also grant that the above three considerations, i.e. co-occurrence, sameness in behaviour, and inference to the best explanation, are at least jointly sufficient for warranting beliefs in a posteriori scientific type-identities like 'water = H₂O', 'heat = mke', and so on. Now, phenomenal-physical type-identities can only be established by way of these considerations if they are just like familiar scientific identities. The question is then: are these scientifically-minded a posteriori physicalists right in thinking that these considerations are also jointly sufficient for establishing a posteriori phenomenal-physical type-identities?

2.3. A DISANALOGY

Let me now elaborate on a number of considerations that seem to favour maintaining a disanalogy here. First, while we have no qualms taking things involved in a scientific identity claim to share the same location, e.g. water and H₂O occupy the same spatiotemporal location, intuitively we do not take conscious experiences and brain states to have the same location. Our folk conception of experience suggests that either experiences are located in specific parts of the body – a toothache is in the tooth, an itch is on the skin of a certain part of the body, etc. – or it does not make sense to think that they have specific locations such as the brain, e.g. experiences like feeling depressed, being in love, and admiring a masterpiece artwork. Questions like 'Where are C-fibres

firing?’ make perfect sense, but questions like ‘Where is it that you feel happy?’ or ‘Where are you experiencing stress?’ hardly seem to make sense.¹²

Second, while scientific identities involve only things that are objective, phenomenal-physical identities involve not only the objective but also the subjective – two essentially different points of view. With respect to a sample of water, we can use electrolysis and spectrography to identify H₂O molecules and work out how water is just H₂O. Now a sample of pain understood as a phenomenal experience is, one might say, a subjective experience – it is private and accessed in a distinctive way by the subject who is having the pain. No matter how much one introspects pain, one is not going to discover that the experience of pain is the firing of C-fibres. Unlike in the case of water, it is also not clear what tools we can use to get to the firing of C-fibres from the experience of pain, because they involve two radically different points of view – one subjective and the other objective.¹³

The main contrast between scientific identity claims such as ‘water is H₂O’ and phenomenal-physical type-identity claims such as ‘the experience of pain in humans is the firing of C-fibres in the brain’, as is indeed acknowledged by physicalists themselves

¹² This is known as ‘the location problem’ for what is known as ‘the identity theory’ first put forward by U. T. Place (1956), Herbert Feigl (1958) and J. J. C. Smart (1959) (see Malcolm 1964; Kim 1966). This problem may not seem insurmountable (see Smart 1959; Kim 1966: 233-4). One might say that while we intuitively treat *pain* and *itches* as locatable objects, experiences like *being in pain* and *being itchy*, understood as events, are located in the brain or where the subject of the experience is located. Identity theorists might also bite the bullet and insist that experiences such as feeling depressed or being in love, which are events, are indeed located in the brain or where the subject of the experience is located.

¹³ A difference in points of view does not entail that the relevant identity claim is false. Identity claims such as ‘Here is Somerville College library’ involve two distinct points of view but are not intuitively false. More on this in the next chapter.

(see Perry 2001; Papineau 2011, forthcoming; Sundström 2017), lies in the following disanalogy. It seems that even if, as most physicalists think, scientists have already told us what pain experience is, we are still left with a sense of bewilderment. We still cannot help but ask questions such as ‘Why is pain experience C-fibres firing?’ or ‘Why is C-fibres firing the experience of pain?’ In contrast, we certainly do not persist in asking why water is H₂O or why H₂O is water. While there is nothing puzzling about identity claims like ‘water is H₂O’, we seem to feel a persistent puzzlement regarding phenomenal-physical type-identity claims.

The above are just a number of considerations one might raise in supporting the claim that phenomenal-physical type-identity claims are disanalogous to scientific identity claims such as ‘water is H₂O’. If one takes these considerations seriously, then one would naturally hold the case of conscious experience to be somewhat unique and different from ordinary physical phenomena. One might take this disanalogy to further show that it is not clear at all that we have strong enough reasons for positing phenomenal-physical type-identities à la a posteriori physicalists.

It is worth noting that the three considerations raised in this section do not just support the claim that *a posteriori* phenomenal-physical type-identity claims are disanalogous to familiar scientific identity claims. The considerations also apply to a priori physicalists who take phenomenal-physical identity to be a priori. These considerations are general considerations against the general identity thesis that physicalists standardly subscribe to. This is the claim that the phenomenal *is* just the physical (including the functional).

2.4. FROM THE PROBLEM OF THE EXPLANATORY GAP TO THE INTUITION OF DUALISM

So far in this section, we have considered two strategies for a reductive explanation of the phenomenal in terms of the physical: one involves conceptual analysis and is favoured by a priori physicalists; and the other involves identity claims and is adopted by a posteriori physicalists. The first strategy seems unsatisfactory because intuitively conscious experiences cannot be given adequate functional analyses in terms of their causal roles. The second strategy also seems unsatisfactory because phenomenal experience seems unique and proposed phenomenal-physical identities seem to be disanalogous to familiar scientific identities. Given the unsatisfactoriness of both strategies, it seems that the problem of the explanatory gap persists and it is questionable that conscious experience can be reductively explained.

However, my aim in this section was not to argue that physicalists have failed to solve the problem of the explanatory gap.¹⁴ It is true that, in discussing their attempts, I have shown that their proposed solutions to the problem of the explanatory gap are intuitively unsatisfactory. But my aim, as stated at the beginning of the section, was rather to use an in-depth discussion of physicalists' attempts at bridging the explanatory

¹⁴ Physicalists might indeed acknowledge the intuition that phenomenal-physical identity claims seem disanalogous to other scientific identity claims. They might then undertake a commitment to give an account of *why* there is such an intuition (see Papineau 2011). They might further argue that once the commitment is taken on and putatively discharged, the three considerations discussed in 2.3 do not really add up to a persuasive case against *a posteriori* type-identity claims or physicalism in general. The question is whether physicalists can satisfyingly explain away the persistent intuition that the experience of pain in humans is not C-fibres firing in the brain. I will address this issue in chapter 3.

gap to bring out the intuition of dualism, the intuition that conscious experience is nonphysical.

In section 2.3, we noted that phenomenal-physical identity claims such as (i) 'the experience of pain (in humans) = the firing of C-fibres in the brain' seem disanalogous to scientific identity claims such as 'water = H₂O'. Importantly, we can't help but keep asking questions such as 'Why is pain experience C-fibres firing?' or 'Why is C-fibres firing the experience of pain?'. We seem to be puzzled by the idea that the phenomenal character of a pain experience, i.e. the painfulness of pain, is just the physical property of being the physical event of C-fibres firing. In contrast, we certainly do not persist in asking why water is H₂O or why H₂O is water. There is nothing puzzling about identity claims like 'water is H₂O'.

This inclination to persist in questioning physical-phenomenal identity claims demonstrates that we have a persistent intuition that conscious experiences are not physical states, that is, phenomenal properties are not physical properties. It seems that we have an intuition of dualism, an intuition that says consciousness is nonphysical.

Many philosophers explicitly acknowledge the intuition of dualism. Some directly assert the existence of the intuition. For instance, Howard Robinson (2015: 119) writes:

The claim that phenomenal red is the same property as electrical activity between neurons, as these are characterized in physical science, is intuitively obviously false in pretty much the same way as it is intuitively obviously false that phenomenal red is the same quality or property as phenomenal blue, or as the sound of middle C.

Others try to elicit the intuition by engaging in certain mental exercises. In the opening section of an unpublished paper, Andrew Melnyk elicits the intuition of dualism by inviting the reader to perform the experiment of gazing at a red object and attending introspectively to the phenomenal property of the experience. Melnyk writes:

Now, while thinking of the property in the way you normally would in introspection – so don't think of it under some description like 'the property I'm now attending to introspectively' – try to take seriously the idea that *that* property literally *is* a physical property, a neural property, perhaps. You will find, I think, that you just can't take this idea seriously – that the idea strikes you as incredible.

The intuition of dualism is also widely acknowledged by avowed physicalists. In a recent paper, Pär Sundström (2017) remarks on the uniqueness of phenomenal-physical identity claims, as discussed above:

Many of us have an exceptional resistance to the physicalist identity thesis. The resistance is exceptional in that we do not have it to other identifications that we have sufficiently good reason to accept, like the identification of liquidity with loose molecular connection, or of Cicero with Tully, or of myself with the shopper who set off the alarm.

Sundström cites physicalists such as John Perry and David Papineau to make the point that even physicalists themselves acknowledge the seeming 'absurdity' of physicalism. According to Perry (2001: 4), as we have already seen in the last chapter, '[t]o say that this, the feeling I am aware of when I, so to speak, look inward, is that, the thing [some brain state] I read about, just seems crazy'.

Papineau's remarks on the intuition of dualism are particularly noteworthy here. Responding to the charge that physicalists have not closed the explanatory gap,

Papineau, in a number of places (2002, 2008, 2011, forthcoming), defends the idea that there is no *problem* of the explanatory gap (and, indeed, no explanatory gap), but only a *feeling* of an explanatory gap which results from an ingrained intuition of dualism.

Papineau (2011: 5) writes:

It is widely agreed among contemporary philosophers of mind that science leaves us with an ‘explanatory gap’ – that even after we know everything that science can tell us about the conscious mind and the brain, their relationship still remains mysterious. I think that this agreed view is quite mistaken. The feeling of an ‘explanatory gap’ arises only because we cannot stop ourselves thinking about the mind-brain relation in a dualistic way.

As an a posteriori physicalist, Papineau thinks that we have sufficient evidence in support of positing a posteriori phenomenal-physical type-identity claims. Thus, there is no explanatory gap – no question of explaining conscious experiences in terms of brain states – because the two are one and the same. Nonetheless, Papineau admits that we still cannot help finding the relationship between conscious experiences and physical states mysterious even though it is not in fact mysterious. It seems that *the intuition of an explanatory gap*, the feeling that the gap has not and cannot be bridged, persists. So Papineau denies the existence of the *explanatory gap* but affirms the existence of the *intuition (or feeling) of an explanatory gap*.¹⁵ The *intuition of an explanatory gap*, argues Papineau (2011: 5), results from the *intuition of dualism*, the fact that ‘we cannot stop ourselves thinking about the mind-body relation in a dualistic way’. Papineau (2011: 11) even goes on to point out that avowed physicalists are in fact closet dualists because of

¹⁵ Note that the *intuition of an explanatory gap* is distinguished from the *explanatory gap*. The existence of the former does not entail that of the latter.

this unshakable intuition.¹⁶ In this sense, one may be persuaded theoretically or philosophically by arguments in favour of the phenomenal-physical identity thesis, but still find dualism somewhat intuitive.

It is worth pointing out here that Papineau appeals to the existence of the intuition of dualism to defend his a posteriori physicalism, because for him, what really underpins anti-physicalists' insistence on the existence of the explanatory gap and resistance to phenomenal-physical identification is just this ingrained dualistic intuition that we have. Now, if the intuition of dualism at issue can be explained in a way that does not pose a problem for the physicalist ontology, Papineau's strategy for defending physicalism would succeed. In that case, we find consciousness unique and refuse to accept phenomenal-physical identities simply because we have an intuition of dualism which does not point to any real duality in nature. In this case, the intuition of dualism does not pose a problem for physicalism. However, in the absence of such an explanation, the intuition of dualism may very well pose a problem for physicalism. For instance, the intuition of dualism could be underpinned by a more basic intuition where the content of this intuition, call it T, is incompatible with physicalism. If physicalists cannot explain away this more basic intuition and cannot defend their position against T, then T poses a problem for physicalism and so does the intuition of dualism.

¹⁶ Papineau (2011) argues that the kind of terminology physicalists employ, e.g. brain processes 'generate', 'yield', 'cause', or 'give rise to' conscious states, suggests that they take phenomenal states to be distinct from brain states. In contrast, as Papineau (2011: 11) writes, 'H₂O doesn't 'generate', 'cause', 'yield' or 'give rise to' water. It is water.'

We will discuss different proposals for explaining the intuition of dualism in the next chapter. In the rest of this chapter, I shall further clarify the intuition of dualism.

3. THE INTUITION OF DUALISM

In this section, we take a closer look at the intuition of dualism. The kind of dualistic intuition under consideration here is also distinguished from the kinds of ‘intuitions of dualism’ discussed in psychology and social-cultural studies.

3.1. DEFINING THE INTUITION OF DUALISM

Following Chalmers (2018: 5), I define the intuition of dualism as the intuition that *consciousness is nonphysical*. There is much debate about what intuitions are, whether they are beliefs, dispositions, *sui generis* states, etc. (Pust 2017). We need not get into this debate here. With respect to our intuition of dualism, we can simply take it to be a belief or disposition to believe that consciousness is nonphysical. This belief or disposition to believe can result in certain of our behaviours, including, notably, making verbal reports and forming judgements such as:

- It is hard to see how consciousness could be physical.
- Phenomenal states are not brain states.
- It is intuitively false that phenomenal properties are physical/functional properties.
- The experience of pain doesn’t seem to be the firing of C-fibres.

As we clarified in chapter 1, terms such as ‘consciousness’, ‘phenomenal states’, ‘phenomenal properties’ and ‘experience’ are predominantly used to mean the same thing in the literature. So while these statements have different contents, their upshot is the same, namely, expressing the intuition *that consciousness – the phenomenal character or phenomenal property of our conscious experience – is nonphysical*.

In addition to having an intuition of dualism, one might also at the same time have an intuition of physicalism. There is nothing incoherent in the idea that a physicalistic intuition and a dualistic intuition can co-exist. The former becomes salient when we consider, for instance, the progress of science, or the ontologically simple picture that physicalism offers. The latter intuition becomes gripping when we contemplate our conscious experience.

Two main characteristics of the intuition of dualism are worth emphasising. First, the intuition of dualism, as our discussion in the last section has shown, seems *persistent*. The possession of the intuition might require one to reflect on one’s conscious experiences. But once the intuition is brought out, it does not go away. Here, I draw a distinction between the intuition of dualism and what is sometimes called ‘the intuition of contingency’. An intuition of contingency is just an intuition about two things, any two things, seeming to be contingently related. For instance, one might have the intuition of contingency that water and H₂O are contingently related. But upon careful reflection, such an intuition might just go away.¹⁷ While the intuition of contingency might not be

¹⁷ As one might say in the manner of Kripke (1980), one is simply confused in thinking that water and H₂O are contingently related – what one takes to be contingently related to H₂O is only something that has the appearance of water, e.g. looking watery, running in rivers, etc., not

persistent, in contrast, as we have seen, the intuition of dualism that consciousness is nonphysical is persistent – it does not disappear even after being told by physicalists that we have good reasons to take consciousness to be physical.

Second, the intuition of dualism also seems to be *widely shared*. It is certainly widely shared among philosophers. One might think that in some cases, the belief or judgement that consciousness is nonphysical is arrived at through a careful consideration of philosophical arguments against physicalism. While this is perfectly possible, here we are only interested in the kind of intuition of dualism that does not result from conscious inferential reasoning through complex philosophical arguments. Physicalists who have considered and rejected philosophical arguments against physicalism can equally share this kind of intuition of dualism. Indeed, as Papineau and others have pointed out, even avowed physicalists are under the spell of the intuition of dualism. One might also suspect that the intuition of dualism is shared by ordinary thinkers who do not have much philosophical training. One might think that so long as one has a clear grasp of the philosophical notion of ‘consciousness’ and some grasp of the notion of ‘physical’, one would have the intuition that consciousness is nonphysical. Of course, in many cases, people will deny having an intuition of dualism. One’s confirmation of the intuition may be trumped by a stronger inclination to, for instance,

water itself. Once one discovers this confusion, one might no longer have the intuition that water is only contingently related to H₂O. Whether or not the intuition of contingency regarding water persists would depend on how we ordinarily think about what water is. If our ordinary non-scientific conception of water is a functional role (causal role) conception, the intuition of contingency regarding water and H₂O might still persist even after being told by scientists that they are identical. We need not take a stance on the issue here. The intuition of contingency, by definition, need not be persistent, and can thus be distinguished from the intuition of dualism, which is persistent.

hold physicalism. But this doesn't mean that there is no underlying intuition of dualism. The precise extent of the intuition among non-philosophers nevertheless remains unclear and can only be determined by empirical means.

3.2. VARIETIES OF INTUITIONS OF DUALISM

Outside philosophy, talk of an 'intuition of dualism' is widespread. Gabriel Segal (2007: 101) notes that the idea of folk dualism, a general mental-physical divide, has featured significantly in both anthropological and psychological literatures. In particular, folk dualism is generally regarded as a cognitive universal in the cognitive sciences (Slingerland and Chudek 2011: 1). The chief advocate of the idea of folk dualism is psychologist Paul Bloom. In his book *Descartes' Baby* (2004), Bloom presents case studies to show that infants automatically develop a dualistic view of the world.

However, it is important to distinguish the kind of mental-physical divide that cultural theorists and developmental psychologists discuss from the kind that we are concerned with.¹⁸ Discussions in psychology and anthropology are dominated by the

¹⁸ The mental-physical divide can be drawn in different ways. In fact, looking at the history of the mind-body problem in western philosophy itself, one would notice that while dualism has been a common theme, different aspects of the mind have been opposed to the body or matter, such that dualism has been understood differently at different times (Robinson 2011). In classical and mediaeval periods, it was the notion of the *intellect* that was contrasted with the physical or the material. For instance, in the *Phaedo*, Plato argues that the intellect is immaterial, because the intellect apprehends the Forms and Forms are immaterial (1997: 68-74, i.e. 78b4-84b8). The mental-physical divide conceived along the lines of a divide between the phenomenal and the physical is often thought to originate from Descartes (Crane and Patterson 2000; Cottingham 2000). This kind of mental-physical divide, i.e. a divide between the phenomenal and the physical, is what lies at the heart of the intuition of dualism which we have been concerned with.

concern with whether we are intuitive *Cartesian substance dualists*. According to this conception of the mental-physical divide, the mind, in the words of Chinese Studies scholar Edward Slingerland (2011: 74), ‘belongs to an ontological realm separate and independent from the realm of the merely physical or bodily’. Similarly, linguist Anna Wierzbicka (2006a: 166; see also 1989) characterises this conception of folk dualism as the idea that people in all cultures think that a person has two parts – one part is the body which is visible and the other part, i.e. the soul, is invisible. Bloom also seems to conceive folk dualism in this way. Bloom thinks that ‘[w]e have two distinct ways of seeing the world: as containing bodies and as containing souls’ (2004: xii), ‘we think of bodies and souls as distinct; we implicitly endorse a strong substance dualism of the sort defended by philosophers like Plato and Descartes’ (2007). However, the idea that there exists a universal folk dualism conceived along the lines of Cartesian substance dualism is often rejected by theorists, especially anthropologists. For instance, medical anthropologist Joseph Alter (1992: 92) writes: ‘In Hindu philosophy the mind and the body are intrinsically linked to one another... There is no sense of simple duality’. As anthropologist Sondra Hausner (2007: 55) puts it, ‘South Asian approaches to the body have been lauded as a theoretical counter to the Cartesian split’.

An intuition of dualism conceived along the lines of Cartesian substance dualism is to be distinguished from the kind of dualistic intuition at issue here. Certainly, when philosophers claim that there is an intuition of dualism, they usually do not mean that substance dualism, such as conceived along the lines suggested by Descartes, is intuitive. It is true that an intuitive Cartesian substance dualist who believes that souls are distinct

from bodies, and believes that only souls, not bodies, have conscious experiences, is likely to believe that conscious experiences are not physical states of the body. Indeed, it may well be the case that for some, the intuition that consciousness is nonphysical is underpinned by an intuition of substance dualism. However, the kind of dualistic intuition that philosophers of mind are primarily concerned with is supposed to hold even if we think that there are no souls that are distinct from bodies. One might also add that the idea of soul conceived as distinct from the body no longer holds an important place in a world-view increasingly dominated by science. In contrast, the kind of dualistic intuition we are concerned with here has not been concomitantly affected by the rise of science. In fact, it is reasonable to suggest that the scientific study of the brain itself has made us more aware of our dualistic intuition. So in discussing what explains our intuition of dualism, I shall only be concerned with the kind of dualistic intuition that is independent from the intuition of Cartesian substance dualism.

Many anthropologists and psychologists, while rejecting a universal Cartesian substance dualism, endorse a weaker conception of universal folk dualism. According to this weaker conception, 'humans are dualistic thinkers in the broad sense – that is, they tend to conceptually sort information from the environment into mutually exclusive categories' (Hodges 2008: 387) – a mental category (which includes thoughts, desires, etc.) and a physical category (which includes physiological functions, habits, etc.) (see Lakoff 1987; Slingerland and Chudek 2011). But it is unclear that this weaker conception of folk dualism is directly related to the kind of dualistic intuition that is at issue here, which contrasts the physical with the phenomenal.

When anthropologists and psychologists talk about how humans tend to think that there is a fundamental divide between the mental and the physical, they don't seem to be contrasting the physical with the phenomenal. Take Bloom's discussion for instance. While it is not completely clear what aspect of the mind is at centre stage in Bloom's conception of folk dualism, in various places, he seems to contrast the physical with the intentional. For instance, writing on art, he thinks that there are two ways of looking at an artwork, which correspond to 'the two ways of seeing the world we possess more generally – in terms of physical bodies and in terms of desires and intentions' (2004: 94). Furthermore, even if Bloom is right in thinking that humans have the tendency to draw a distinction between the physical and the intentional, this sheds no light on why we have an intuition of dualism to the effect that the phenomenal character of experience is nonphysical. A tendency to draw a division between intentional agents and physical bodies does not mean that there is a *persistent* intuition of dualism pertaining to the *intentional* and the physical, let alone a persistent intuition of dualism pertaining to the *phenomenal* and the physical.

In conclusion, while there is much talk about folk dualism outside philosophy, it is unclear that this literature bears direct relevance to the kind of intuition of dualism that we are concerned with here, and which is made particularly salient by our discussion on the problem of the explanatory gap in the context of the metaphysical debate about the nature of consciousness.

4. CHAPTER CONCLUSION

Our discussion on the problem of the explanatory gap brings out the counterintuitiveness of phenomenal-physical type-identity claims. We saw that identity claims like ‘the experience of pain in humans = the firing of C-fibres in the brain’ seem to be disanalogous to identity claims such as ‘water = H₂O’. We cannot help but ask questions such as ‘Why is the experience of pain in humans just the firing of C-fibres in the brain?’ In contrast, we do not persist in asking ‘Why is water H₂O?’ This disanalogy shows that there exists a persistent intuition of dualism, an intuition that consciousness is nonphysical. We also saw that the intuition of dualism at issue here is not to be confused with an intuition of Cartesian substance dualism, or the human tendency to distinguish the mental from the physical. The dualistic intuition at issue pertains specifically to consciousness or the phenomenal. Such an intuition is commonly acknowledged by philosophers including physicalists, and is plausibly widely shared by non-philosophers.

Having teased out the relevant intuition of dualism pertaining to consciousness, and how it is to be distinguished from other kinds of intuitions of dualism, in chapter 3 we turn to proposals for explaining the intuition of dualism.

CHAPTER 3

EXPLAINING THE INTUITION OF DUALISM: SOME PROPOSALS

1. INTRODUCTION

In the last chapter, we concluded that there exists a deep, shared and persistent intuition of dualism with respect to consciousness, an intuition that consciousness is nonphysical. Given that there exists such an intuition, what explains it? This chapter examines a number of candidate explanations.

The structure of the chapter is as follows. Section 2 clarifies what counts as an adequate explanation of the dualistic intuition. Section 3 looks at an account from Fiala, Arico and Nichols (2011) which explains the intuition of dualism by appealing to dual-process cognitive architecture. Section 4 examines a proposal found in Nagel's book *The View from Nowhere* (1986) according to which the subjective and the objective do not seem to converge to a common reality. Section 5 discusses an account put forward by Papineau (2002) which explains the intuition of dualism in terms of the uniqueness of phenomenal concepts, concepts we employ in thinking about conscious experiences. I argue that these proposals all fail to be satisfactory. Section 6 concludes Part I and introduces Part II.

2. ADEQUATE EXPLANATION

The task at hand, namely, explaining the intuition of dualism, requires us to address what counts as an adequate explanation in this case. Often, what counts as an adequate

explanation is context-dependent. Crucially, it depends on the audience whom the explanation is for and the particular explanatory concerns of this audience. An adequate explanation thus addresses such explanatory concerns. Consider the question: 'Why does salt dissolve in water?' If this question is raised in a primary school classroom, then the explanation – 'salt is soluble and soluble things dissolve in water' – is likely to count as a perfectly good explanation. But such an explanation is unlikely to be satisfying or even appropriate if the question is raised in a high school chemistry class. In the latter case, the person who raises the question is likely to be interested in an explanation at the microscopic level about how components of salt interact with components of water (e.g. salt is made of positive sodium ions and negative chloride ions, and the positive part of water molecules attracts the former and the negative part attracts the latter.) The proposed explanation simply does not address this and thus fails to be adequate.

Let's turn to our explanandum – the intuition of dualism. As we have clarified in the last chapter, the intuition of dualism is the belief or the disposition to believe that consciousness is nonphysical. So the relevant question we are asking is: 'Why is it the case that we believe or are disposed to believe that consciousness is nonphysical?'

Someone who raises this question may be interested in the structure of our belief system, what it is that underpins our intuition of dualism. We treat people's beliefs and actions as governed by reasons. Similarly, one might expect our intuition of dualism, our belief or disposition to believe that consciousness is nonphysical, to be intelligible or to exhibit a certain kind of reasonableness given our other beliefs and attitudes. In this case, our explanatory concern in asking what explains the intuition of dualism resembles that

of someone who raises the question – ‘Why do people in certain collectivist cultures believe that family members have moral responsibilities to look after the elderly?’ – with the intention of understanding the beliefs and value system of people in those cultures. Given this kind of explanatory interest, the resulting explanation for the intuition of dualism is a *rational explanation* – an explanation that uncovers the underlying rationale for our dualistic intuition. Such an explanation may appeal to other beliefs or attitudes we hold.

Rational explanations are certainly abundant in our ordinary explanatory practices. In everyday contexts, we often explain a person’s attitudes and actions in terms of that person’s beliefs and desires, where these latter beliefs and desires may very well be unconscious to the subject. Similarly, to say that our intuition of dualism is underpinned by a certain rationale does not mean that those who have the intuition arrive at the intuition through conscious reasoning. One can hold a belief *p* or have a disposition to believe *p* without being consciously aware of one’s own reasons for believing *p* or being disposed to believe *p*.¹

In asking for an explanation of our intuition of dualism, one might alternatively be interested in an explanation about how our cognitive mechanisms and underlying

¹ Philosophers distinguish between *motivating reasons* and *explanatory reasons* (e.g. Alvarez 2010; Dietz 2016). Motivating reasons are ‘reasons that motivate someone to ϕ ’, and explanatory reasons are ‘reasons that explain why someone ϕ -s or ϕ -ed’ (Alvarez 2010: 39). Motivating reasons form a subcategory of explanatory reasons; not all explanatory reasons are motivating reasons. The reason that explains why John volunteers every weekend may be that John is an altruistic person. But that John is an altruistic person is not a motivating reason for John to volunteer every weekend. A motivating reason features as a premise in the subject’s reasoning, either theoretical or practical, about ϕ -ing. Such reasoning, as Alvarez (2010) notes, may be explicit or implicit. If the reasoning is implicit, the subject is not consciously aware of the reason that motivates her to ϕ .

processes give rise to our belief or tendency to believe that consciousness is nonphysical. In such a case, one is interested in a *mechanistic explanation*, an explanation at the level of machines or systems.

We can say that a mechanistic explanation is a *subpersonal-level explanation*, whereas a rational explanation is a *personal-level explanation*. The distinction between personal-level explanation and subpersonal-level explanation was first introduced by Daniel Dennett (1969). The relevant distinction between levels is a distinction in terms of different kinds of descriptions.² Personal-level explanations refer to persons as wholes, and are cashed out in terms of beliefs, sensations, desires, etc. of persons. Subpersonal-level explanations do not make reference to persons as intentional agents but focus on the physical processes underpinning the attitudes and actions of persons, and are cashed out in the vocabulary of cognitive science or neuroscience.

Apart from rational and mechanistic explanations, there are also other kinds of explanations of a commonly held belief or a disposition to form such a belief. Consider the religious beliefs that there exists a God, that there is an afterlife, etc. Karl Marx (1970), for instance, provides a *historical/cultural explanation* for our religious beliefs, according to which such beliefs are responses to the alienation and oppression in the material world and products of indoctrination imposed by the ruling class over the working class. Jesse Bering (2006) offers an *evolutionary explanation* for people's belief in an afterlife,

² The terminology 'personal' and 'subpersonal' may suggest that the relevant distinction is a distinction between the level of wholes and the level of parts. As a number of philosophers have pointed out, the relevant distinction here should not be understood in part/whole terms (See Davies 2000; Hornsby 2000).

according to which such a belief results from evolutionary adaptation possibly aimed at promoting altruistic behaviour. Explanations of these kinds might also be possible with respect to the intuition of dualism.

In asking for an explanation of our intuition of dualism, we are leaving open our explanatory concern or interest, and are thus not looking for a specific kind of explanation. So, we are open to different kinds of explanations here. Importantly, we need not think of different explanations as mutually exclusive or necessarily competing with each other. For instance, given the personal/subpersonal distinction, we can have explanations for the same phenomenon at different levels of description. Our intuition of dualism may be explained at the subpersonal level in terms of cognitive processes, and at the personal level in terms of reasons. Such an explanation could, one might think, be purely empirical, in which case it would be up to psychologists to provide the whole story. Alternatively, it might be possible to explain the intuition of dualism satisfactorily through careful armchair reflection about the special way we think about conscious experience. Psychologists might then be able to test the armchair explanation empirically. The task of explaining the intuition of dualism could turn out to be an interdisciplinary project.

So far, we have seen that an adequate explanation for the intuition of dualism can potentially take various forms and addresses the explanatory concern of the audience whom the explanation is for. I shall now clarify some further desiderata that such an explanation must satisfy.

An adequate explanation for the intuition of dualism, first of all, must address the target phenomenon, in this case, our intuition of dualism. If a proposal explains some related phenomenon but fails to explain the intuition at issue, it automatically fails to be satisfactory. An adequate explanation must also have predictive powers, predicting correctly the range of cases where the intuition of dualism arises. We are made consciously aware of our dualistic intuition upon being confronted with certain cases or scenarios, which further result in our making the relevant judgements or verbal reports. For instance, as discussed in the last chapter, in considering the phenomenal-physical identity claim ‘the experience of pain (in humans) = the firing of C-fibres in the brain’, we have the persistent inclination to judge that the identity claim is intuitively false. This then is a case where the intuition of dualism arises, in the sense that the intuition is made evident or conscious to the subject. An adequate explanation for the intuition of dualism should at least predict correctly with respect to these cases.

Consider then the following desiderata for an adequate explanation of the intuition of dualism (provided that the explanation addresses the target intuition):

- (a) The explanation should not *undergeneralise*: in cases where an intuition of dualism or a similar intuition of distinctness would arise, the explanation must predict accordingly.
- (b) The explanation should not *overgeneralise*: in cases where no intuition of dualism or no similar intuition of distinctness arises, the explanation must not predict that it would arise.

According to the two desiderata above, an adequate explanation of the intuition of dualism ought to predict occurrences of the intuition of dualism as well as similar intuitions of distinctness. The intuition of dualism, as we saw, is the intuition that phenomenal consciousness is nonphysical. One might think that it is not only the phenomenal that seems intuitively distinct from the physical. A similar *intuition of distinctness* regarding the physical and some other putatively nonphysical phenomenon might occur in a different domain. Here, one might wonder what a similar intuition of distinctness might be. In chapter 7, I shall consider whether there is a similar intuition of distinctness regarding physical properties and colour properties.³

Given the two desiderata for assessing the adequacy of an explanation for the intuition of dualism, let us now move on to examine proposals for explaining this intuition. In his paper ‘The Meta-Problem of Consciousness’, Chalmers (2018) surveys a wide range of proposals that are aimed at explaining what he calls ‘problem intuitions’ about consciousness. As we saw in chapter 1, the intuition of dualism is one of the problem intuitions. Many of the proposals that Chalmers has surveyed are problematic or unsatisfying for obvious reasons. I shall not repeat Chalmers’ assessments here. In the following, I focus on three proposals in detail. The first, ‘dual-process cognition’, and the

³ Some might say that there is also a similar intuition of distinctness that intentionality is nonphysical. I think if there is such an intuition of distinctness, it may be a product of our dualistic intuition regarding consciousness. The thought here is that some intentional states are associated with phenomenology or phenomenal properties, and the intuition of distinctness regarding intentional states and physical states hinges on the presence of phenomenology (see Papineau 2011: 15). I also do not think that there is a similar intuition of distinctness of moral properties from physical or natural properties. While it is often claimed that our commonsense intuitions favour moral realism over moral anti-realism (e.g. Sayre-McCord 2015), there does not seem to be a strong intuition which says that moral properties are non-natural or nonphysical properties.

second, 'the irreducibility of subjectivity', are not explicitly addressed by Chalmers,⁴ and the third, 'the antipathetic fallacy', is addressed by Chalmers (though not under this title), but is worthy of in-depth discussion here. Assessments of all these three proposals bring out clearly what needs to be explained in explaining the intuition of dualism at issue. I shall argue that these proposals either fail to explain the target intuition of dualism or fail to satisfy the desiderata.

3. DUAL-PROCESS COGNITION

Brian Fiala, Adam Arico and Shaun Nichols (2011) put forward a detailed account of the origin of our dualistic intuition by appealing to a specific feature of our cognitive architecture, namely, dual-process cognition. They argue that there are two cognitive pathways via which we attribute conscious states to others. On the one hand, there is the high-road pathway, which is slow and produces conscious-state attributions through our rational deliberation, theory application, or conscious reasoning. On the other hand, according to Fiala et al, there is also the *low-road* mechanism that is fast, operates on a restricted range of inputs, and is not under conscious control. The low-road mechanism to conscious-state attribution is described by what they call the 'agency model', according to which once we categorise an entity as an *agent* (i.e. as Fiala et al. put it, representing the entity as having any of the three features: appears to have eyes; appears to behave in a contingently interactive manner; or displays distinctive non-inertial

⁴ With respect to 'dual-process cognition', Chalmers (2018: 25) mentions it and claims that he is not sympathetic to this proposal without spelling out exact reasons.

motion trajectories), we are disposed to attribute consciousness to the entity (2011: 92). So for Fiala et al., attributions of consciousness are closely related to attributions of agency – if we categorise the entity as an agent, then we would naturally attribute consciousness to it, and if we do not so categorise it, we would not attribute consciousness to it.⁵

Fiala et al. argue that the dual-process cognition model of conscious-state attributions provides an explanation for our dualistic intuition.⁶ Their explanation of the intuition is as follows: Through high-road reasoning about the brain, e.g. reasoning about the plausibility of physicalism, we can arrive at, so they claim, the conclusion that ‘specific brain processes or brain regions *are* conscious experiences’ (2011: 96). But in this case, the low-road mechanism is not activated – when considering the lump of matter that is the brain, all features which might trigger in us an agency categorisation are lacking. So the low-road mechanism fails to trigger attributions of the corresponding conscious experiences to the brain. This, according to Fiala et al., ‘leave[s] us feeling as if something has been left out: our low-level, *low-road* process remains silent where it would normally provide intuitive confirmation of our *high-road* output’ (2011: 97). So, on

⁵ According to Fiala et al., the two cognitive pathways often converge, i.e. produce the same output, but not always. They give the example of conscious-state attributions to insects: in this case, the low-road mechanism is triggered, leading to the attribution of consciousness to insects; but the high-road mechanism does not yield the same attribution – we typically reject the claim that insects have consciousness after taking the limitations in their neural systems into consideration.

⁶ Fiala et al. do not in fact use the terminology ‘intuition of dualism’. They use terms such as ‘an explanatory gap intuition’, ‘an intuition of a gap’ or ‘a “gappy” intuition’ interchangeably. However, they do explicitly claim that their account ‘provides an important part of the explanation for why dualism is so attractive’ (2011: 88). Here I shall assume that their account is also aimed at addressing our intuition of dualism.

their account, the intuition of dualism occurs when high-road reasoning focuses on the brain as the seat of consciousness, but the brain does *not* trigger the low-road mechanism.

This psychological approach to explaining the intuition of dualism does not suggest a duality in nature between phenomenal properties and physical properties and is thus perfectly compatible with physicalism. Given the two desiderata laid out in section 2 regarding an adequate explanation of the intuition of dualism, one might wonder whether this explanation would overgeneralise since it relies on the presence of dual-process cognitive architecture and dual-process cognition may be implicated in many domains. Indeed, given this account, we should expect an intuition of something being left out if the high-road mechanism is deployed in a case that does not engage the low-road mechanism. Fiala et al. (2011: 103) suspect that similar intuitions can occur in domains such as intentionality and causation. For instance, with respect to causation, they contend that our intuition that counterfactual theories of causation leave out the ‘oomph’ of causation is similar to our intuition of dualism regarding the case of consciousness. They suggest that there could be two pathways in the case of causal attributions. Our low-road mechanism is triggered by perceiving changes taking place after objects have been in contact. So our ‘gut-level sense of causation seems to be driven by a low-road system that is insensitive to covariation information’, and *covariance* is the kind of notion found in high-road descriptions (2011: 103), such as philosophical theorising about causation. According to Fiala et al. (2011: 104), our intuition that reductive explanations of causation, e.g. counterfactual theories of causation framed in terms of the notion of covariance, are unsatisfying might be explained by ‘the failure of

such explanations to trigger the low-road processes that generate the gut-level sense that A caused B'.

In both the case of consciousness and that of causation, the high-road pathway engages with information, about the brain and about covariation, which does not trigger the low-road system, and therefore does not lead to the relevant attributions from the low road, e.g. X is conscious, A causes B. This lack of activation of the low-road mechanism then gives rise to the respective intuitions of distinctness. As Fiala et al. would say, just as intuitively we do not take conscious states to be brain states, intuitively we do not take causation to be counterfactual covariance.

But one might wonder whether the intuition that causation is not just covariance is as persistent and widely-shared as the intuition that consciousness is nonphysical. *Prima facie*, we should be cautious about treating the two cases similarly.

It seems that Fiala et al. group these cases together because of their questionable assumption about what needs to be explained in explaining our dualistic intuition regarding consciousness. Underlying their proposal is the implicit characterisation of the intuition of dualism in terms of *attributions*, primarily *third-person* attributions, of consciousness – we intuitively attribute consciousness to humans and animals, while finding it incredible that consciousness is physical. For instance, they write:

People are happy to credit consciousness to cats, but it is counterintuitive that cat-consciousness is ultimately nothing more than populations of neurons firing synchronously at 40–60Hz. That is where our proposal enters the picture. (Fiala, et al. 2011: 102)

According to their account, the intuition of dualism arises because while we attribute consciousness to the brain via the high-road pathway, the brain does not trigger the low-road system of consciousness attribution which is underlined by what they call the 'agency model'.

But this way of understanding the intuition of dualism ignores our phenomenal conception of conscious experience. One might think that central cases where the intuition of dualism arises are cases where *I* am having a conscious experience and reflect on my experience. Note that in such a case, my first-person attribution that *I am conscious* is not underpinned by the agency model, based on a prior categorization of myself as an agent. In such a case, there is an intuition of dualism – as I reflect on my experience, I become gripped by the compelling intuition that having certain neurophysiological features in my brain is not all there is to me when I have my conscious experience.⁷

Indeed, the intuition that conscious experiences are not brain states does not depend on whether or not we ascribe consciousness or agency to the brain, as Fiala et al. seem to think. The intuition is simply that conscious experiences, phenomenally conceived, just don't seem to be brain states. To conceive a conscious experience phenomenally is to consider it in terms of its phenomenal character or its what-it-is-likeness for the subject as the subject undergoes the experience. Our intuition of dualism is thus the intuition that the phenomenal character of our experience is not physical.

⁷ Fiala et al. would say that in cases of reflecting on one's own experiences, we also have an intuition of dualism because when we think of our own brain, we do not categorise it as an agent – we do not think of it as something that has eyes, behaves and moves in a certain way, etc. (see Fiala et al. 2011: 102).

Intuitively, such an intuition can be made salient regardless of whether or not we are attributing consciousness or agency to the relevant entities.

So, Fiala et al.'s proposal is problematic because it is mainly concerned with *third-person attributions* of consciousness and what facilitates such attributions, i.e. dual-process cognition. It is also problematic because it assimilates attributions of consciousness to attributions of agency. Consequently, their characterisation of the intuition of dualism neglects our phenomenal conception of conscious experience and fails to capture what is crucial to the intuition. In conclusion, their proposal for explaining our intuition of dualism fails to be satisfactory, not exactly in the sense of failing to meet the two desiderata laid out in section 2, but because it fails to explain the right intuition.

4. THE IRREDUCIBILITY OF SUBJECTIVITY

In chapter 2 section 2.3, we considered the idea that phenomenal-physical type-identity claims are disanalogous to ordinary scientific identity claims because our conscious experience involves the subjective point of view, whereas physical phenomena are objective. The idea of the subjective point of view featured notably in Thomas Nagel's 1986 book *The View from Nowhere*. According to Nagel (1986), subjectivity and objectivity mark two different points of view – an inner perspective versus an outer perspective. Persons are subjects who can think about the world, and their own thoughts and actions, from a particular point of view, i.e. 'the subjective point of view'. Persons can also

transcend this point of view and think about everything in a detached manner, i.e. from an objective point of view.

Given Nagel's idea, one might wonder if our dualistic intuition can be explained in the following way. There can be two conceptions of things or properties in the world: a subjective conception and an objective conception. Our conception of conscious experience is a subjective conception, whereas our conception of a physical state is an objective conception. There is no a priori route from the objective to the subjective. This a priori separation of the two kinds of conceptions, one might say, has the consequence that we cannot understand how the two conceptions converge on a common reality. Hence, we are left with an intuition of dualism.

Is this a good explanation for our intuition of dualism? It seems not. According to the explanation, there is something deeply mysterious about how subjective and objective conceptions can converge on a common reality, and this mystery is supposed to explain our intuition of dualism. But the convergence of the subjective and the objective on a common reality is not always mysterious. For instance, when Tom thinks, 'I am Tom', the subjective conception (the 'I' conception that he has of himself) and the objective conception (a public 'Tom' conception) converge on the same person. There does not seem to be any particular mystery surrounding how that could be the case. In this case, we do not have a problem endorsing the identity statement 'I am Tom' when uttered by Tom. Nor do we have qualms endorsing other identity claims which also involve two distinct points of view, e.g. 'Here is Somerville College, the first college in Oxford which admitted women', when uttered by someone who is standing in the

courtyard of Somerville College; 'Now it is noon GMT on the 15th of February 2018', uttered at noon GMT on the 15th of February 2018, etc.

The examples mentioned above involve indexical terms, i.e. 'I', 'here' and 'now'. Indexical conceptions involve a subjective point of view. There is no a priori entailment from propositions about people, places, and times described in context-independent ways to the respective propositions about a person as 'I', or a place as 'here', or a time as 'now'. The failure of a priori entailment holds in these indexical cases, just as in phenomenal cases. However, while it does not seem mysterious that the subjective and the objective in these indexical cases can converge on the same reality, it does seem mysterious how the phenomenal, which is subjective, and the physical, which is objective, can converge on the same reality as physicalists describe. It seems that some part of a satisfactory explanation of our dualistic intuition must distinguish between the phenomenal on the one hand, and the indexical or the merely subjective on the other, although Nagel groups them together as the subjective. So, appealing to the distinctness of the subjective from the objective, on its own, does not explain why the phenomenal and the physical are intuitively distinct in reality.

It is worth mentioning that Nagel (1986: 31) himself takes points of view, the subjective and the objective, to be irreducible features of reality. But this point, as we have just seen, is highly controversial. In writing a review for Nagel's 1986 book, Christopher Peacocke (1989) criticises Nagel for making an unjustified 'sense/reference jump'. This is the move Nagel makes from asserting that there are distinct, irreducible

standpoints or forms of understanding to asserting that there are distinct, irreducible features of reality.

In summary, we do think that the kind of subjectivity that pertains to phenomenal experience is ontologically irreducible to the objective physical order. But our intuition of dualism is the way it is, not solely in virtue of the distinction between subjectivity and objectivity. Thus, this explanation does not adequately explain our intuition of dualism. It fails to satisfy desideratum (b) – in cases where no similar intuition of distinctness arises, it predicts that it would arise.

5. THE ANTIPATHETIC FALLACY

I now turn to an explanation for the dualistic intuition proposed by Papineau (2002).⁸ Papineau (2011: 15) thinks, and I agree, that ‘the intuition of dualism hinges on the way we think specifically about phenomenal states’. Phenomenal states are thought of with phenomenal concepts. Physical states are thought of with physical concepts. For Papineau, the intuition of dualism prototypically arises when encountering phenomenal-physical identity claims such as ‘the experience of pain is the firing of C-fibres’. These identity claims seem intuitively false, and they involve both phenomenal concepts and physical concepts. So it would seem that if we can explain why

⁸ This account was put forward by Papineau (2002) in his book *Thinking about Consciousness*. In a later paper, Papineau (2011) discusses and acknowledges some of the problems regarding his account.

phenomenal and physical concepts do not seem to co-refer, we have an explanation for the intuition of dualism.

Papineau's proposed explanation for the intuition of dualism appeals to certain apparent differences between phenomenal concepts and physical concepts. Phenomenal concepts are first and foremost a priori distinct from physical concepts. As we have seen in the last section, this a priori distinctness also holds regarding indexical and non-indexical concepts, and is not sufficient for explaining our intuition of dualism. Something more needs to be said about phenomenal concepts to distinguish them from indexical concepts. According to Papineau, the exercise of a phenomenal concept is 'accompanied by an actual or imagined instance of the phenomenal state being referred to' (2011: 16). But this phenomenology is lacking when we exercise the corresponding physical concept. On the basis of the latter disparity between phenomenal concepts and physical concepts, Papineau puts forward an explanation for the intuition of dualism. Consider the following phenomenal-physical identity claim:

The experience of pain (in humans) = the firing of C-fibres

According to Papineau, when we entertain this identity claim, our thinking about the left-hand side is accompanied by the experience of pain itself or a faint imagined copy of it; whereas thinking about the firing of C-fibres on the right-hand side involves no such experience. This, according to Papineau, gives one the impression that the right-hand side fails to equate with the left-hand side because it leaves out the crucial element of phenomenology. Consequently, we find phenomenal-physical identity claims mysterious.

Papineau's proposed explanation for the intuition of dualism is supposed to help defend physicalism, which he advocates. Given this explanation for the intuition that phenomenal-physical identity claims seem false, Papineau contends that we should not deny these identities. He (2002: 170-1) argues for this by attending to the structural differences between phenomenal and physical concepts and pointing to an analogy with the distinction between *use* and *mention*. Experiences themselves are *used* or *present* in our thinking about experiences via phenomenal concepts. This is the reason why in exercising a phenomenal concept, one feels like one is having the experience itself (2002: 170). Physical concepts 'leave out' experiences only insofar as the exercise of physical concepts does not *use* or *activate* experiences. This does not mean that physical concepts do not *mention* or *refer* to experiences, or that physicalism is false. To conclude that physicalism is false would be to commit a fallacy, which Papineau terms 'the antipathetic fallacy'.

However, there are some problems with Papineau's account. First, one may question Papineau's idea that the exercise of a phenomenal concept is always 'accompanied by an actual or imagined instance of the phenomenal state being referred to' (2011: 16). The claim here is that in employing a phenomenal concept of experience E, one either has the experience E (i.e. in introspecting E) or an experience that phenomenally resembles E (e.g. in imagining or remembering E), i.e. 'an imagined instance'. Tim Crane (2005: 160) has raised the worry that there can be non-introspective cases where one can exercise a phenomenal concept without having any experience resembling the experience being referred to. For instance, I have the nonphysical,

phenomenal concept of what it is like to taste durian fruit, because I had durians in the past; I could recognise the taste if I were to taste one right now; and I can also describe the taste to some extent – it tastes creamy, not very sweet. However, I do not call to my mind the taste whenever I think of durians – this could simply be because I am very bad at imagining tastes in general. So it seems that my ordinary thoughts about durians, which employ the phenomenal concept of the taste of durian, *contra* what Papineau claims, are not accompanied by ‘imagined instances’ of the experience of tasting durians.

In response, as Crane (2005: 161) points out, Papineau could say that the concept I have of the taste of durian, call it C, is in fact *not* a phenomenal concept. So there are then three different kinds of concepts: physical concepts, phenomenal concepts and concepts like C (which are not physical concepts; but also are not quite phenomenal concepts because the exercise of these concepts, like the exercise of physical concepts, does not necessarily involve having any experience). Papineau’s account only covers instances of the dualistic intuition arising when we exercise physical concepts and the corresponding phenomenal concepts as he defines them in the special sense. But one might think that there is also an intuition of dualism when we exercise a concept like C and then go on to consider the corresponding physical concept. If so, then Papineau’s explanation does not satisfy desideratum (a) mentioned in section 2, because in certain cases where a dualistic intuition does arise, it fails to predict accordingly. It seems that Papineau would have to deny that there is an intuition of dualism in these cases.

But even if we grant that Papineau can deny that the intuition of dualism arises in the aforementioned cases, his proposal faces other problems. Sundström (2008) has

argued that Papineau's proposal falsely predicts dualistic intuitions in cases where there are none. Consider the following example from Sundström (2008: 141):

(S) The experience of off-white = my brother's most salient
current experience

According to Papineau's account, the left-hand side is conceived imagistically (i.e. is accompanied by a faint copy of seeing off-white) and the right-hand side is not, since I do not know what my brother's most salient current experience is. So the account predicts the occurrence of the dualistic intuition with respect to (S), and that (S) is intuitively false. But this certainly does not seem to be the case – in fact we have no idea whether (S) is true or false. When considering this identity claim, I might not agree with this identification, because I do not have a clue what my brother's most salient current experience is. But my doubt about the identity is due to a lack of evidence for the co-reference of the two concepts (Sundström 2008: 142). It has nothing to do with the intuition of dualism. We don't think that (S) is intuitively false. It seems that Papineau account incorrectly predicts the occurrence of the intuition of dualism in (S) when there is none, thus failing to satisfy desideratum (b).

While there is no intuition of dualism or a deep counterintuitiveness involved in taking 'the experience of off-white' to be identical with 'my brother's most salient current experience', in contrast, there is an intuition of dualism regarding the identity claim 'my brother's most salient current experience = such-and-such neural activity in the visual cortex'. With respect to the latter, it seems that the exercise of neither the concept on the left-hand side nor that on the right-hand side is accompanied by an experience. Given

the latter case, Papineau's account fails to satisfy desideratum (a) – it predicts that there is no intuition of dualism where there seems to be one. Given also the counterexample of (S), it seems that Papineau's account fails to meet both desiderata (a) and (b).

In response to Sundström's counterexample (S), Papineau (2011) thinks that there is room for manoeuvre in his account. According to Papineau, it is not clear that the right-hand side does not activate an experience. He (2011: 17) suggests that maybe we tend to surreptitiously activate an experience on the right-hand, for instance, filling out the reference of 'my brother's salient current experience' by imagining the experience of seeing off-white. Alternatively, he also suggests that in cases like (S), we tend not to activate the experience on the left-hand side.⁹ However, Papineau does not argue for those claims, which, as a result, makes his response unsatisfying and somewhat *ad hoc*. Indeed, one could simply multiply examples to problematise Papineau's account (see Kammerer 2018a).¹⁰ What we need to see from Papineau is a clarification of the

⁹ It is not enough that either (i) we tend to activate an experience on the right-hand side, or (ii) we tend not to activate an experience on the left-hand side. For the response to work, Papineau needs to say that either (i') we *always* activate an experience on the right-hand side, or (ii') we *never* activate an experience on the left-hand side. For instance, given (i), if there is one single occasion when an experience is not activated by the right-hand side, then there would be an intuition of dualism, and so a counterexample to the account.

¹⁰ In a recently published paper, François Kammerer (2018a) constructs thought-experiments and elicits counterexamples like the following to further problematise Papineau's proposal:

- (i) My experience of this blue square = the state of my brain when it detects this blue square
- (ii) The current state of Elise's visual cortex = the state of Elise's visual cortex when she looks at this red oval

Let's consider (i) first. According to Kammerer, we can stipulate that when I think of the left-hand side, I introspectively focus on my experience of a particular blue square that I am looking at, and that when I think of the right-hand side, I think of the 'blue square' component on the basis of my current visual experience of that particular blue square. Similarly, regarding (ii), we can stipulate that I think of 'this red oval' on the basis of my visual perception of it. So in (i),

conditions under which an experience is activated with respect to the employment of a concept. Without such clear criteria, there could always be counterexamples like (S).

Furthermore, Papineau's response also seems to be damaging to his position. If the response is that we tend to activate the experience of seeing off-white on the right-hand side, in which case (S) would not seem intuitively false, there is a question as to just why the experience of seeing off-white is activated. One might think that it is only because the identity claim (S) is being accepted as true. But this leads to the following worry. If the original phenomenal-physical identity, i.e. 'The experience of pain (in humans) = the firing of C-fibres', is accepted as true, then, by parity of reasoning, the deployment of the concept of 'C-fibres firing' on the right-hand side might lead to activation of the experience of pain. There would then be no intuition of dualism given Papineau's proposal. But there is precisely an intuition of dualism even for physicalists who accept phenomenal-physical identity claims. Alternatively, if the response is that we tend not to activate an experience of seeing off-white on the left-hand side of (S), then one might worry that in cases like the original identity claim, we might also tend not to activate the experience of pain on the left-hand side. If so, then there is simply no intuition of dualism with respect to the original phenomenal-physical claim. But again, Papineau's starting point is precisely that the intuition of dualism occurs when we consider phenomenal-physical identity claims. So, the hypothesis that there might be

phenomenal concepts are exercised on both sides of the identity claim, and in (ii) only the right-hand side involves the exercise of a phenomenal concept. Intuitively, there is an intuition of dualism in (i) but not in (ii). But Papineau's account incorrectly predicts that there is no intuition of dualism in (i) and there is one in (ii).

surreptitious activation or non-activation of experience seems damaging to Papineau's own position.

Overall, it does not seem that Papineau's proposal gets to the bottom of our dualistic intuition. The above two problems raised against Papineau demonstrate that our dualistic intuition is tied to there being a crucial difference between, on the one hand, typical phenomenal concepts, e.g. the experience of off-white, or concepts which clearly refer to conscious experiences, e.g. my brother's most salient current experience, my concept of the taste of durian, etc., and, on the other hand, physical concepts, e.g. such-and-such neural activity in the visual cortex. Concepts of the former kind track the presence of phenomenal properties, and concepts of the latter kind track the presence of physical properties. But the intuitive distinctness of the two kinds of properties, tracked by the two kinds of concepts as taxonomised above, does not seem to be adequately accounted for by Papineau's proposal.

6. CONCLUSION OF PART I

Part I now comes to an end. In chapter 2, we saw that there exists a deep, shared and persistent intuition of dualism, an intuition that says that consciousness is nonphysical. This chapter clarified what an adequate explanation of the dualistic intuition might look like, and critically assessed three candidate explanations of the intuition. As we saw, none of the proposals are satisfactory. I do not claim that the proposed explanations for

the intuition of dualism surveyed in this chapter exhaust all possible explanations.¹¹ This, however, should not pose a problem for the explanation proposed in Part II. The explanation I put forward is a personal-level rational explanation. There might very well be other evolutionary or mechanistic explanations for the intuition of dualism at issue.

In Part II, which spans four chapters, I propose a different explanation of the intuition of dualism. This requires us to revisit a familiar idea in the philosophy of mind, namely, *revelation*, the claim that the nature of phenomenal properties is revealed in experience. My ultimate aim in Part II, especially in chapter 7, is to make clear and plausible the idea that our intuition of dualism is crucially due to our tacit or implicit understanding of the thesis of revelation, which, as I argue, is part of our ordinary conception of experience.

¹¹ For criticisms of some of the other proposals, see Chalmers (2018).

CHAPTER 4

REVELATION

1. INTRODUCTION

In analytic philosophy, the term ‘revelation’, introduced by Mark Johnston (1992), is used in the debate about the metaphysics of colour to refer to the view that the essence of colour is revealed in colour experience. (I will say more about the thesis of revelation in the colour debate in chapter 7.) The parallel thesis of revelation in the contemporary mind-body debate is sometimes put as the claim that the essence of experience is revealed in experience (see Stoljar 2009). This latter thesis of revelation is by no means unfamiliar to philosophers of mind, although the name ‘revelation’ has not always been used to identify it. Discussions on revelation can be traced to Descartes and Hume. Descartes (1985: 215, 217) thought that upon a careful reflection, we are capable of having *clear and distinct ideas* of experiences such as sensations. To have a clear and distinct idea of something, according to the standard interpretation of Descartes (Gewirth 1943; Patterson 2008), is to know, at least to some extent, what belongs and does not belong to the *essence* of that thing.¹ One might also read Hume as advocating a version of revelation.

¹ We can think of the essence of x as constituted by a collection of essential properties $\{E_1, E_2, E_3, \dots E_n\}$. (I will say more about the notion of essence later.) According to the standard interpretation of Descartes, one has a clear idea of x if and only if one knows that some particular essential property or properties belong to the essence of x . So one’s idea of x can become clearer as one comes to know more essential properties of x . One has a distinct idea of x if and only if it is not the case that one takes some property other than $E_1, E_2, E_3, \dots E_n$ to belong to the essence of x .

Hume thought that impressions, which are roughly feelings, including ‘sensations, passions and emotions’ (T 1.1.1.1), ‘appear, all of them, in their true colours’ and ‘must necessarily appear in every particular what they are, and be what they appear’ (T 1.4.2.7; see also T 1.4.2.5, T 2.2.6.2).²

One can also find contemporary discussions on revelation in relation to the general topic of conscious experience and more specifically the physicalist-dualist debate about the nature of consciousness (Strawson 1989; Lewis 1995; Stoljar 2006, 2009, 2018; Nida-Rümelin 2007; Damjanovic 2012; Montague 2016; Goff 2015, 2017; Majeed 2017; Trogdon 2017). In the contemporary literature, the thesis of revelation is commonly understood as a claim about knowing the phenomenal properties of conscious experience. It is this sense of revelation that is relevant to the debate on the metaphysics of consciousness and that I shall be focused on expounding in this chapter. Notably, this sense of the thesis of revelation was explicitly formulated in David Lewis’ 1995 paper ‘Should a Materialist Believe in Qualia?’. Lewis takes revelation (or what he calls ‘the

According to Descartes (1985: 215), we can have *clear and distinct ideas* of the mind and the body, just as we can have *clear and distinct ideas* of their principal attributes (i.e. their essences), thought and extension. For Descartes, nothing physical or bodily belongs to the essence of the mental. Descartes (1985: 208) also thought that our ideas of *sensations*, e.g. having pain, seeing colours, etc., although ‘indeed very clear’, are ‘not always distinct’. Nevertheless, he maintains that our ideas of sensations are capable of becoming distinct through thoughtful reflection. Descartes (1985: 217) notes, ‘pain and colour and so on are clearly and distinctly perceived when they are regarded merely as sensations or thoughts’.

² All references to Hume are in this style of citation, i.e. book, part, section, paragraph. ‘T’ stands for ‘*A Treatise of Human Nature*’ (1739-40). How precisely to understand the passages cited here is a point of contention pertaining to the interpretation of Hume’s doctrine of *mental transparency*, according to which we cannot be mistaken with respect to (certain aspects of) our mental states, and we cannot fail to be aware of (certain aspects of) our mental states. For a survey of different interpretations of Hume’s doctrine of *mental transparency*, see Qu (2017). Here I am merely suggesting that attributing a version of the thesis of revelation to Hume is at least a plausible interpretation of Hume.

Identification Thesis') to be incompatible with physicalism, which he is committed to, but at the same time finds the thesis highly intuitive. More recently, Philip Goff (2015: 214) defined revelation as the claim according to which 'a psychologically normal subject can come to know the real nature of one of her phenomenal qualities by attending to that quality'. Goff (2011, 2015, 2017) used revelation to argue against physicalism.

This chapter takes a close look at the thesis of revelation. The incompatibility between revelation and physicalism will be spelt out in the next chapter. The structure of the chapter is as follows. Section 2 contains a preliminary clarification of the notion of essence which features in the thesis of revelation. Section 3 examines Lewis's remarks about revelation, on the basis of which I put forward a precise formulation of the thesis in section 4. Section 5 and 6 clarify two important points with respect to revelation. In section 5, I make clear that the thesis of revelation involves a knowledge-attribution *de dicto* regarding knowledge *de dicto* that a particular quale Q is X, where the predicate 'X' captures the essence of Q. In section 6, I clarify how the thesis of revelation might plausibly entail the claim that (some) qualia are simple. Section 7 sums up the chapter.

2. ON ESSENCE

Revelation is a claim about knowing the essences or natures of qualia.³ Before we turn to the precise formulation of the thesis of revelation, it is worth saying something about how I understand the notion of *essence* first. There are two main lines of thought

³ 'Essence' and 'nature' are used interchangeably in this context.

regarding the notion – one understands essence in modal terms; and the other takes the notion of essence to be primitive and unanalysable, conceiving it on the model of definition. In his famous 1994 paper ‘Essence and Modality’, Kit Fine argues against the previously dominant *modal account* of essence, which analyses the notion of essence in terms of modality, i.e. an essential property of a thing is just a necessary property of that thing, and the essence of a thing is a property that is both necessary and sufficient for being that thing. Fine provides several counterexamples to the modal account. For instance, Socrates has the necessary property of belonging to the singleton set that has Socrates as its only member, i.e. {Socrates}. But intuitively, it is not part of the essence of Socrates that he belongs to the singleton set {Socrates}. Fine takes examples like this to show that the modal account of essence is false – an essential property of a thing is not simply a necessary property of that thing.

Fine, for his part, opts for the *real definitional* account of essence, which understands the notion of essence on the model of definition. According to this account, as Fine (1994: 2) puts it, ‘the notion of definition has application to both words and objects – ... just as we may define a word, or say what it means, so we may define an object, or say what it is’. This account of essence traces back to Aristotle. According to Aristotle in *Metaphysics* (1031a12), ‘definition is the formula of the essence’; the definition of a thing states the essence of that thing, i.e. what it is to be that thing.⁴ This account is

⁴ The Ancient Greek notion at issue is *to ti ên einai* (‘the what it was to be’) or sometimes *to ti esti* for short (‘the what it is’), which was rendered as ‘essentia’ by Roman translators, from which we get our English word ‘essence’ (Cohen 2016).

also found in Locke (1689/1996: 185), who takes the essence of a thing to be ‘the being of anything, whereby it is, what it is’.⁵

In the discussion of the thesis of revelation, the notion of *essence* will be understood in the Aristotelian/Lockean/Finean sense, i.e. ‘the what it is’ – that which makes a thing the thing it is – which is stated by a definition.⁶ The definition of a thing states what it is to be that thing, what that thing is (essentially). Fine (1995b) discusses two ways of expressing an essentialist claim: by using a predicate modifier, as in ‘Water is essentially ...’; and by using a sentential modifier, as in ‘It is true in virtue of the identity of water that water is ...’. Here I shall be using the former, i.e. ‘X is essentially ...’, to make claims about essence. For instance, consider what it is to be water. To be water is to be molecules made up of two hydrogen atoms and one oxygen. Water is, by definition and essentially, molecules made up of two hydrogen atoms and one oxygen atom. Consider what it is to be a sister. To be a sister is to be a female sibling. Sisters are, by definition and essentially, female siblings. Now consider what it is to be the phenomenal property of pain experience, i.e. the painfulness of pain, the phenomenal property that is instantiated by all and only experience-tokens that belong to the pain

⁵ The notion of essence is meant to reside with Locke’s notion of ‘real essence’ rather than his notion of ‘nominal essence’.

⁶ There is the question of whether the essence of x is identical to or ontologically distinct from x . In discussing Aristotle’s notion of essence, Anna Marmodoro (2009) points out that Aristotle, having put forward the question of ‘whether each thing and its essence are the same or different’ in *Metaphysics* VII 6, goes on to take a thing to be ‘one and the same as its essence’. Here I will adopt the assumption that x is one and the same as the essence of x , e.g. a phenomenal property is one and the same as its essence. This assumption does not affect the crucial points I make with respect to revelation.

experience-type.⁷ If to be that phenomenal property is to be the physical property of being an event of C-fibres firing, then the phenomenal property of pain experience is, by definition and essentially, the physical property of being an event of C-fibres firing. Such a definition states the putative essence of the phenomenal property of pain experience, what it is to be that property.

It is also worth clarifying different senses of the notion of 'essence' in comparison to the relevant sense at issue. Throughout the dissertation and especially in formulating the thesis of revelation, by 'essence', I mean that which belongs to the thing's *most core respects*. This is what Fine (1995a: 276) calls 'immediate constitutive essence' (see also Fine 1995b; Dasgupta 2014: 589). *Constitutive* essence is contrasted with *consequential* essence (Fine 1995a: 276). If it is essential to Q that X then it is *consequentially* essential to Q that XvY. *Immediate* essence is contrasted with *mediate* essence (Fine 1995a: 281). Consider this example to bring out the relevant distinction here: *being a sister* is essentially and by definition *being a female sibling*. *Being a female sibling* is the immediate essence, as well as the constitutive essence, of *being a sister*.⁸ Now, *being a sibling* is essentially and by definition *being one of the two or more individuals with one common parent*. *Being one of the two or more individuals with a common parent* is the immediate constitutive essence of *being a sibling*. So, *being female and being one of the two or more individuals with one common parent* is, in Fine's terminology, the *mediate* essence of *being a sister*. The

⁷ Although in his original paper Fine (1994) was mainly concerned with individual-essence, one can extend the Finean conception of essence to apply to type-essence.

⁸ For Fine (1995a: 276) '[t]he constitutive essence of an object, strictly conceived, is its immediate constitutive essence'. So the distinction between constitutive and consequential essence cuts across the distinction between immediate and mediate essence. The notion of immediate essence is one aspect of the notion of constitutive essence (Fine 1995a: 282).

Aristotelian/Lockean/Finean notion of essence that is commonly appealed to in discussion (Fine 1995a: 276; Dasgupta 2015: 589), and which is what I am appealing to here in formulating the thesis of revelation, refers to *immediate constitutive* essence, *not* mediate nor consequential essence.

In the following section, I shall take a closer look at Lewis' insightful discussion on the thesis of revelation. This will help to bring out the formulation of revelation proposed in section 4.

3. LEWIS ON REVELATION

In his article 'Should a Materialist Believe in Qualia?' Lewis discusses the thesis of revelation, which he calls 'the Identification Thesis'. Lewis' account of revelation is framed in terms of the notion of *qualia*. The term 'qualia' (plural; *quale*, singular) in this context refers to phenomenal properties of particular events of experiencing. According to Lewis, although 'quale' and 'qualia' are technical terms, the concept is nevertheless part of folk psychology. As a functionalist, Lewis regards the term 'qualia' as 'a name for the occupants of a certain functional role that is spelled out in our tacitly known folk psychology' (1995: 140). So qualia are properties of experiences and they play a certain functional role that is characterised folk-psychologically. According to Lewis (1995: 141), the folk-psychological role of qualia includes the following components:

- (i) 'Causing abilities to recognize and to imagine experiences of the same type.'

- (ii) Being 'responsible for responses of pleasure, disgust, etc.'.
- (iii) Being 'responsible for judgements of similarity-distance, e.g., the judgement that these two colour-samples nearly match whereas those two contrast strikingly'.

These three components of the qualia-role seem to be platitudinous. But according to Lewis (1995: 141), 'there is more to the folk-psychological concept of qualia'. Lewis introduces the fourth component:

- (iv) 'We *identify* the qualia of our experiences.'

Lewis elaborates on this component of the qualia-role:

Folk psychology says, I think, that we *identify* the qualia of our experiences. We know exactly what they are – and that in an uncommonly demanding and literal sense of 'knowing what'. If I have an experience with quale Q, I know that I am having an experience with quale Q, and I will afterwards remember (unless I happen to forget) that on that occasion I had an experience with quale Q. It is by producing this identifying knowledge that a novel experience confers abilities to recognize and imagine. (1995: 141)

Consider the experience of tasting the Australian spread Vegemite. In having such an experience, one might have various thoughts about the experience and come to know numerous things about it, e.g. 'This experience is interesting', 'I have had this experience before', 'It is similar to the experience of tasting the British spread Marmite', etc. Among these thoughts is, Lewis would say, the following one: 'I (now) know exactly what the phenomenal character of the experience of tasting Vegemite is'. This latter knowledge is an instance of what Lewis calls 'the Identification Thesis' – 'We know exactly what [qualia] are' (1995: 141). According to Lewis, one's abilities to recognise and imagine a

certain type of experience are explained by the Identification Thesis – we know exactly what the qualia of our experiences are and such identifying knowledge confers these abilities.

But what does Lewis mean when he says that ‘we know exactly what [qualia] are’? Lewis clarifies:

I spoke of ‘an uncommonly demanding and literal sense of ‘knowing what’.’ Let me elaborate. 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: 142)

According to Lewis, by having an experience with a particular quale, we are in a position to know exactly what that quale is in the sense that we are in a position to know the essence of the quale. This is what Lewis calls ‘an uncommonly demanding and literal sense of “knowing what”’. This demanding sense of ‘knowing what a quale is’ contrasts with a less demanding sense of ‘knowing what’ (1995: 143). In everyday contexts, we can say that one knows what something is even if one does not know the essence of that thing. One knows what water is in an everyday non-scientific context insofar as one knows a cluster of descriptions about water, and knows how to identify ordinary samples of water, etc. To know what water is in this less demanding everyday sense of ‘knowing what’ does not require one to know the essence of water, i.e. know that water is H₂O.

4. THE FORMULATION

What Lewis calls ‘the Identification Thesis’ is just the thesis of revelation (see Stoljar 2006, 2009). As we have just seen, Lewis (1995: 142) formulates the thesis of revelation as the following:

[W]hen I have an experience-token with quale Q, the knowledge I thereby [am in a position to] gain reveals the essence of Q: a property of Q such that, necessarily, Q has it and nothing else does.

There are two things to notice in Lewis’ formulation of the thesis of revelation. First, Lewis is adopting the simple modal account of essence, according to which the essence of something is that which is both necessary and sufficient for being that thing. We have already seen the well-known problem associated with this account of essence as discussed by Fine (1994). Here, we can simply take the notion of essence in the intuitive Aristotelian/Lockean/Finean sense, conceived on the model of definition, which was laid out in section 2.

Second, on the face of it, Lewis takes the essence of Q, where Q is phenomenal property, to be a second-order property. To know exactly what Q is, for Lewis, is to know that phenomenal property Q has the (second-order) property X where X is the essence of Q (that is, Q has X and nothing else does). Without resorting to talk of second-order properties, here we can simply understand ‘knowing exactly what a quale is’, i.e. knowing the essence of a quale, as knowing some truth.⁹ (This makes the task of

⁹ Here knowledge-what is understood as a kind of knowledge-that. This is the orthodox approach to knowledge-what in the literature (see Stanley & Williamson 2001). Knowing what *x* is at least involves knowing a proposition which is the answer to the question ‘What is *x*?’.

formulating the thesis of revelation simpler, and is in fact what Lewis goes on to say.¹⁰) We can say that one knows *exactly* what the property P is if and only if one knows *that* 'P is thus and so' where 'thus and so' *defines* P, that is, captures the essence of P.¹¹ For instance, one knows exactly what the property *being triangular* is if and only if one knows a truth that captures what the property (by definition and essentially) is, e.g. 'Being triangular is having a three-sided closed shape'. The whole proposition is a definition of the property of *being triangular*, stating what it is to be triangular. The predicate 'having a three-sided closed shape' captures the essence of triangularity. In the same way, knowing exactly what quale Q is amounts to knowing the truth 'Q is X', where 'X' is a predicate that captures Q's essence. The thesis of revelation says that one is in a position to know this truth about Q merely on the basis of having an experience with quale Q.

However, in the case of a quale, if the thesis of revelation is true, one might think that the truth 'Q is X' is hard to put into words. It seems that with many experiences, we find ourselves devoid of words to describe the qualia or the phenomenal characters of these experiences. Let us call the characteristic quale (phenomenal property) of

¹⁰ Immediately after the passage quoted above, Lewis writes:

If, for instance, Q is essentially the physical property of being an event of C-firing, and if I identify the qualia of my experience in the appropriate 'demanding and literal' sense, I come to know that what is going on in me is an event of C-firing. Contrapositively: if I identify the quale of my experience in the appropriate sense, and yet know nothing of the firing of my neurons, then the quale of my experience cannot have been essentially the property of being an event of C-firing.

Note that *being an event of C-firing* is, strictly speaking, *not* a property of the quale Q which is the qualitative character of pain-experience, i.e. quale *pain*. The property *being an event of C-firing* is more appropriately understood as a property of pain-experience-token itself.

¹¹ Again, *essence* here is understood on the model of definition – the essence of a thing is what is stated by the definition of that thing.

experiences of red things, 'phenomenal red'. If, as I have or recall an experience with that quale, I attempt to define what the quale phenomenal red is, it might seem that all I am able to say is that 'It is *that*', using a demonstrative to refer to the phenomenal character of the phenomenal red experience. But intuitively, the demonstrative 'that' is merely a placeholder for the rich understanding of phenomenal red I have which I am unable to put into words.¹² The fact that I cannot put my understanding into words does not mean that I don't know exactly what phenomenal red is.

Given our discussion, we can sharpen Lewis' idea of revelation as well as our preliminary formulation, and formulate the thesis in the following way:

(Revelation)

By having an experience-token with phenomenal property Q, one is in a position to know a truth, namely, 'Q is X', where the predicate 'X' captures the essence of Q, although it may be hard to put into words.

There are several points worth clarifying here. First, the knowledge we have with respect to quale Q is a kind of knowledge-that or knowledge *de dicto*. The dictum is 'Q is X'. More needs to be said about this knowledge *de dicto*. I will discuss this in the next section.

¹² This demonstrative way of defining phenomenal character does not communicate successfully the content of the utterer's knowledge. This is because demonstratives require ostension, and the hearer might not have cognitive access to what the speaker is thinking of or 'pointing' at. Elizabeth Camp (2006) points out that instead of using demonstratives, the speaker can appeal to metaphors, which do not require ostension, in order to communicate her intended content.

Second, the knowledge *de dicto* 'Q is X' is not just any truth about Q; it is a statement about what makes Q the property it is – the predicate 'X' *captures* the essence of Q. More needs to be said about the notion of 'capture' at issue. I will address this in section 6 when discussing Lewis' argument from the thesis of revelation to the simplicity of qualia.

Third, according to the above formulation, the thesis of revelation does not imply that whenever one undergoes an experience with quale Q, one automatically knows the essence of Q; it only implies that one is *in a position* to know the essence. Being in a position to know in this case requires one to have the kind of cognitive architecture that enables a subject to form *de dicto* thoughts about the phenomenal properties of their experiences. Once this is in place, actually coming to know the relevant essence-capturing truth 'Q is X' is not particularly demanding. All it takes is for one to attend to the experience, not to be distracted, etc. Indeed, the thesis of revelation is supposed to capture the idea that we are often in the right circumstances to actually know that 'Q is X' where the predicate 'X' describes accurately the essence of Q.

There is one more caveat worth pointing out. In order to know, by having an experience with phenomenal property Q, the truth 'Q is X', where 'X' captures the essence of Q, one first needs to know that the experience has the phenomenal property Q. That is, being in a position to know that 'Q is X' requires one to first be in a position to introspect correctly that the experience has Q. In the literature on introspection, doubts have been raised regarding whether we infallibly or even reliably identify the phenomenal character or phenomenal property of our experience (e.g. Schwitzgebel

2011). One might further contend that there are cases such that in having an experience with phenomenal property Q, one is not even *in a position* to introspect that the experience has Q (Schwitzgebel 2011: 129). For instance, one might say that in seeing a hen with 48 clear speckles on its facing side, my experience has the phenomenal character of a particular 48-speckled-ness, which differs from the phenomenal character of seeing a hen with 47 clear speckles on its facing side.¹³ But in having the experience of seeing a 48-speckled hen, it seems that I am not *in a position* to discern this particular 48-speckled-ness of my experience. As a normal human being, I simply lack the introspective capacity to know that my experience has the phenomenal property of 48-speckled-ness, or to tell it apart from an experience with the phenomenal character of 47-speckled-ness.¹⁴ If cases of this kind are plausible, then in such a case, in having an experience with phenomenal character Q, one would not be in a position to know that the experience has Q, and hence not be in a position to know that 'Q is X', where the predicate 'X' captures the essence of Q. That is, the thesis of revelation, as formulated above, would be rendered false. In order to accommodate these problematic cases, we can modify the above formulation of revelation as the following:

(Revelation*)

By having an experience-token with phenomenal property Q *and*
having correctly identified that the experience-token has Q, one is in a

¹³ Whether this is true is debatable. But for the purpose of discussion, let's grant that there is a phenomenal difference here.

¹⁴ It would not help to say that I can know by simply counting the number of speckles on the hen. Counting the speckles would be an altogether different experience from the experience of seeing a 48-speckled hen at a glance.

position to know a truth, namely, 'Q is X', where the predicate 'X' captures the essence of Q, although it may be hard to put into words.

(Revelation*) will be our official formulation. However, for ease of writing, in the following discussion, I will work with the formulation labelled '(Revelation)'. This will not affect the relevant points being made in what is to follow.

The proposed formulation of the thesis of revelation, as we saw, builds on the work of Lewis (1995). It is worth comparing this formulation with some other existing formulations in the literature:

- (1) 'phenomenal concepts reveal the nature of the phenomenal properties they refer to' (Trogon 2017: 2345).
- (2) 'phenomenal concepts reveal the complete nature of the conscious states they refer to' (Goff 2017: 124).
- (3) 'we know essential truths about our qualia on the basis of introspection' (Majeed 2017: 86).
- (4) 'the nature of consciousness is revealed to us in introspection' (Chalmers 2016: 190).

I have formulated revelation as a thesis about the essences or natures of qualia or phenomenal properties. On the first impression, not all of the above formulations of revelation are about the essences or natures of qualia. Some formulations do not explicitly mention 'qualia' or 'phenomenal properties'; they instead mention 'conscious

states' (Goff 2017: 124) and 'consciousness' (Chalmers 2016: 190). However, by these latter phrases, philosophers of mind often simply mean 'qualia' or 'phenomenal properties'. The hard problem of consciousness is a problem about qualia, about how physical processes in the brain can give rise to the phenomenal characters of our experiences (Chalmers 1996). The term 'conscious states' is often used to refer to types of conscious states (see Goff 2017: 107), and the latter are often understood as phenomenal properties (see Sundström 2011: 268).

I have taken revelation to be a thesis about knowing the essence, that is the *immediate constitutive* essence, of a quale in having an experience with that quale. It is unclear that in the above formulations, these philosophers understand the relevant notion of essence or nature in the narrow sense, i.e. in terms of 'immediate constitutive essence'. For instance, terms like 'essential truths' and 'complete nature' might be used to indicate a broad conception of *essence* that includes more than *immediate constitutive essence*.

The above formulations also seem to suggest that these authors are fundamentally interested in a common nature shared by all phenomenal properties, since the singular 'nature' is used as in 'the nature of phenomenal properties', 'the complete nature of conscious states', and 'the nature of consciousness'. According to the formulation I have put forward here, the thesis of revelation is not a claim that the common nature of phenomenal properties is revealed by experience; it is a claim that the (immediate constitutive) essence of *a quale* is revealed by an experience with that quale.

Furthermore, the above formulations characterise the thesis of revelation with respect to either the notion of *phenomenal concepts*, i.e. (1) and (2), or the notion of *introspection*, i.e. (3) and (4). Although my formulation does not make explicit reference to introspection or phenomenal concepts, it can alternatively be formulated with respect to these notions.

On some accounts, our rich understanding of phenomenal properties is embodied in demonstrative-like, or recognition-based, phenomenal concepts (e.g. Chalmers 2010b). An advocate of revelation would say that phenomenal concepts are special in the sense that if one has a phenomenal concept *C* which refers to *x*, the possession of the concept affords one knowledge about, or puts one in a position to know, the essence of *x*. So, given this understanding of phenomenal concepts, the thesis of revelation can thus be alternatively formulated in terms of phenomenal concepts.¹⁵

Introspection is the means by which one can, in a way that no one else can, directly acquire knowledge about one's occurrent or very recently past mental state simply by having the state. Introspection is certainly closely related to the thesis of revelation since the latter makes a claim about what one knows about the essences of qualia by having phenomenal experiences. Presumably, having correctly identified that one's experience-token has phenomenal property *Q* on the basis of having an experience with *Q*, which would then put one in a position to know the essence of *Q*, requires one to have *introspected* the phenomenal properties of the experience. So plausibly the thesis

¹⁵ I did not opt to formulate revelation in terms of the notion of phenomenal concepts because the latter terminology introduces specific theories of phenomenal concepts. It would thus be helpful to sidestep this terminology in the basic formulation of revelation.

of revelation entails that introspection puts one in a position to know the essences of phenomenal properties of experiences. The relevant notion of introspection should be understood in a non-committal and inclusive way without implying any theoretical claims about the nature of introspection or that of self-knowledge.

5. KNOWLEDGE *DE DICTO* VERSUS KNOWLEDGE *DE RE*

In the last section, we have seen that the thesis of revelation is about a kind of knowledge *de dicto*. This point is worth dwelling on. In this section, I shall clarify that the thesis of revelation is best understood as involving a knowledge-attribution *de dicto* regarding knowledge *de dicto* that a particular quale Q is X, where the predicate 'X' captures the essence of Q. To aid the discussion, I appeal to the following two distinctions (see Broackes 1986):

(1) *knowledge de dicto* vs. *knowledge de re*;

(2) *knowledge-attribution de dicto* vs. *knowledge-attribution de re*

(1) is a distinction between different kinds of knowledge. Knowledge *de dicto* consists in a relation between the knower and a certain proposition, whereas knowledge *de re* consists in a relation between the knower and a thing (see Sosa 1970: 883).¹⁶ (2) is a distinction between different types of reports of knowledge. Knowledge-attributions

¹⁶ Knowledge *de dicto* and knowledge *de re* are not mutually exclusive. For instance, I know that the sun will rise tomorrow. I not only have knowledge with respect to a *dictum*, namely, 'The sun will rise tomorrow', I also have knowledge with respect to a *res*, namely, the sun.

de re, but not knowledge-attributions *de dicto*, permit substitution of co-designating terms *salva veritate* (i.e. preserving the truth-value of the proposition).

Given the two distinctions above, one must not assume that knowledge *de dicto* is always reported with knowledge-attributions *de dicto*, or knowledge *de re* reported with knowledge-attributions *de re*. Which kind of knowledge-attribution to use often depends on the context in which the knowledge report is made, to whom the report is made, the purpose of the report, etc. If a thinker thinks about something, she thinks about it under some mode of presentation, e.g. thinks of Venus *as Hesperus*, or *as Phosphorus*. But a report of the thinker's thinking might not specify that mode of presentation. In general, a knowledge-attribution *de dicto* specifies the mode of presentation of the subject's knowledge, whereas a knowledge-attribution *de re* does not. As such, if we are interested in the mode of presentation, we should report the knowledge at issue with an attribution *de dicto*.

For instance, suppose Tom thinks of Hesperus *as Hesperus* and has knowledge *de dicto* that Hesperus appears in the evening sky. Suppose also that Tom does *not* think of Hesperus *as Venus*. In such a case, it would still be correct to report Tom's knowledge in the following way: 'Tom knows, *of Venus*, that it appears in the evening sky'.¹⁷ But this attribution *de re* does not specify the way Tom thinks about Venus and consequently is silent about whether or not Tom knows *de dicto* that Hesperus appears in the evening

¹⁷ In having knowledge *de dicto* that 'Hesperus appears in the evening sky', Tom also has knowledge *de re* of Hesperus/Venus. Such knowledge *de re* can also be reported with a knowledge-attribution *de dicto* – in this case, namely, 'Tom knows *that* Hesperus appears in the evening sky'.

sky. It seems that Tom's knowledge is best reported in the *de dicto* way – 'best reported' in the sense that the report is as accurate and complete a guide as possible with respect to the knowledge at issue, with respect to the way in which Tom thinks about Venus (Broackes 1986: 375).

With the two distinctions above, let us return to the topic of revelation. Revelation says that by having an experience with phenomenal property Q, one is in a position to know a truth or a *dictum*, namely, 'Q is X', where the predicate 'X' captures the essence of Q. It seems that this knowledge *de dicto* is accurately reported by a knowledge-attribution *de dicto* as below:

(D) By having an experience with quale Q, S is in a position to
know *that* Q is X, where the predicate 'X' captures the
essence of Q.

To report it with a knowledge-attribution *de re* fails to completely capture the kind of knowledge at issue. A knowledge-attribution *de re* in this case looks something like this:

(R) By having an experience with quale Q, S is in a position to
know *of* X, where 'X' captures the essence of Q, that Q is it.

Compare (R) with the *de re* report, 'Tom knows *of* Venus that Hesperus is it'. Suppose Tom thinks of Venus *as Phosphorus* and knows *de dicto* that Hesperus is Phosphorus. In such a case, the *de re* report is true. But this does not imply that Tom knows *de dicto* that Hesperus *is* Venus. The *de re* report is compatible with Tom having no idea what Venus is. Equally, (R), while stating that S is in a position to know *de re* of X, that Q is it (where the predicate 'X' captures the essence of quale Q), does not imply that S is in a position

to know *de dicto* that Q is X (where the predicate ' X ' captures the essence of quale Q). In general, one can have knowledge *de re* of the essence, E , of something, Φ , without knowing *de dicto* that Φ is E . Let's agree that water is H_2O , where the expression ' H_2O ' captures the essence of water. Someone who has no knowledge of chemistry may still know *de re* of H_2O – thought of *as water* – that water is it, although she does not know *de dicto* that water is H_2O . Similarly, there could be cases of knowledge *de re* of X that Q is it *without* knowledge *de dicto* that Q is X . In such cases of knowledge *de re*, X is thought of under some mode of presentation but *not as* X . So, (R) is compatible, in a perfectly good sense, with S having no idea of what X is, and with S having no knowledge *de dicto* that Q is X . In the next chapter where we cash out the incompatibility between revelation and physicalism, we will see that for a physicalist, ' X ', which captures the essence of Q , is a physical or functional predicate, and while a physicalist can endorse (R), she cannot endorse (D).

To sum up, revelation is *not* the claim that experience merely puts one in a position to know *de re* of X , under some mode of presentation – not necessarily *as* X – that Q is it (where the predicate ' X ' captures the essence of quale Q) (see also Stoljar 2009). Rather, revelation is the claim that experience puts one in a position to know *de dicto* that Q is X (where the predicate ' X ' captures the essence of Q). This knowledge *de dicto* is best reported with a knowledge-attribution *de dicto* in the manner of (D). The knowledge *de dicto* that ' Q is X ' may not be such that it can readily be expressed in words, but it does involve a rich understanding of X , which is what Q essentially is.

6. REVELATION AND THE SIMPLICITY OF QUALIA

Another notion that is also worth pausing on is the notion of *capture* in the formulation of the thesis of revelation which I have presented in section 4. Recall the formulation:

(Revelation)

By having an experience-token with phenomenal property Q, one is in a position to know a truth, namely, 'Q is X', where the predicate 'X' captures the essence of Q, although it may be hard to put into words.

The word 'capture' has a number of meanings. 'Capture' here should *not* be taken just to mean 'refer to, latch onto'. It is meant to convey the idea of 'representing accurately in words or pictures' as in 'A vivid photograph *captures* the scene'.¹⁸ For instance, the truth 'Being triangular is having a three-sided closed shape' captures or describes accurately what the property of being triangular is. Note that this truth describes the

¹⁸ According to Brian Loar (1997), the word 'capture' is in fact ambiguous. Loar distinguishes two different senses of the claim that 'a concept captures the essence of the referent'. He writes:

These are equivocating uses of 'capture the essence of'. On one use, it expresses a referential notion that comes to no more than 'directly rigidly designate'. On the other, it means something like 'be conceptually interderivable with some theoretical predicate that reveals the internal structure of' the designated property. But the first does not imply the second. (Loar 1997: 603)

For Loar, who is an a posteriori physicalist, a phenomenal concept can be said to *capture* the essence of the phenomenal property in the sense that it directly refers to the referent, but it does not tell us anything about the nature of the referent. This is not how I have intended to use the term 'capture'. On the other hand, a physicalist like Loar would say that a physical concept *captures* the essence of the phenomenal property, in the sense that it tells us what the property really is in a way that reveals the nature of the property. This is the sense of 'capture' at issue here. The thesis of revelation says that the truth 'Q is X', which one is in a position to know through having an experience of the type Q and which is hard to put into words, is a truth that captures, in this second sense, the essence of Q.

internal structure of the property of triangularity – having three sides and being closed. If a property P has an internal structure, assuming that a property's structure is essential to that property (i.e. part of the immediate constitutive essence of that property), then a truth that captures the essence of property P must also describe the internal structure of P. If a truth that captures the essence of P does not describe P as having an internal structure, then P does not have an internal structure, i.e. it is *simple*.

Recall that in our formulation of the thesis of revelation, the truth 'Q is X' is supposed to describe Q's essence. In cashing out revelation, we agreed that 'X' might be difficult (even impossible) to put into words. In cases where Q seems indescribable, we might just say 'this' (pointing inwardly to an experience of Q we have in mind). The intuition that at least in some cases we cannot say more about Q's essence than 'Q is *this*' arises from the fact that in these cases, in having an experience with quale Q, we do not seem to notice Q to have any structure. Given revelation and assuming again that a property's structure is essential to that property, if quale Q had internal structure, then by having an experience with Q, one would be in a position to know that Q had that structure. The description of that structure would be given in stating that Q is X, where the predicate 'X' captures the essence of quale Q.¹⁹ If there is no structure to notice with respect to Q, then Q has no structure, i.e. Q is simple. Indeed, in the following passage,

¹⁹ It seems that whenever we notice our phenomenal character to have structure, we can describe the phenomenal character, including its internal structure, to some extent which goes beyond using a mere demonstrative 'this'. We might say things like: 'There is a bit of phenomenal red on the left of my visual field. Next to it, there is some phenomenal blue, etc.' From this, it does not follow that a complex, structured experience can be fully captured in a verbal description. The simple constituent phenomenal properties will likely not be captured in words without resorting to demonstratives and/or metaphors.

Lewis remarks on the simplicity of qualia, which he takes to follow from the thesis of revelation (1995: 142, fn 4):

If we know exactly what the qualia of our experiences are, they can have no essential hidden structure – no ‘grain’ – of which we remain ignorant. (If we didn’t know whether their hidden ‘grain’ ran this way or that, we wouldn’t know *exactly* what they were. Whatever we might know about them, we would not fully know their essence.) But if nothing essential about the qualia is hidden, then if they seem simple, they *are* simple. We may assume that if a property is structural, then it is so essentially. Then it is a consequence of the Identification Thesis that if we fail to notice structure, there is no structure there to notice. But we do fail to notice structure. So the simplicity of the qualia is a consequence of the Identification Thesis (*inter alia*), and so a derivative part of the folk-psychological concept of qualia.

The above passage can be formulated into the following argument:

- (1) Structure is part of the essence of a property.
- (2) By having an experience-token with phenomenal property Q, one is in a position to know a truth, namely, ‘Q is X’, where the predicate ‘X’ captures the essence of Q.
- (3) If Q has internal structure, then by having an experience with Q, one would be in a position to know that Q has internal structure.
- (4) There is at least a range of cases such that in having an experience with Q, one is not in a position to know that Q has internal structure.
- (5) Some qualia have no internal structure (i.e. they are simple).

Premise (1) seems intuitive. If a property has a certain structure, it has it essentially – a property lacking that structure would not be the same property. (2) is the thesis of revelation. (3) follows from (1) and (2). (4) seems intuitively true – in having a phenomenal red experience, hearing middle C, etc., we just do not notice any structure in the phenomenal characters of these experiences. (Some other qualia do seem to have structural features, e.g. the quale of a certain complex visual experience.) Conclusion (5) then follows from (3) and (4). So for Lewis, the thesis of revelation also entails that (some) qualia are simple (see also Adams 1995 on the simplicity of qualia).

7. CHAPTER CONCLUSION

Drawing on Lewis' ideas in his paper 'Should a Materialist Believe in Qualia?' (1995), in this chapter we formulated the thesis of revelation in terms of knowledge *de dicto* of an essence-capturing truth about a particular quale. We saw that this knowledge, which may be hard to put into words, is best reported with a knowledge-attribution *de dicto*: by having an experience-token with phenomenal property Q, one is in a position to know *de dicto* that 'Q is X', where the predicate 'X' captures the essence of Q. We also saw that the idea that (some) qualia are simple plausibly follows from the thesis of revelation. In the next chapter, we will turn to the incompatibility between revelation and physicalism.

Chapter 5

REVELATION AND PHYSICALISM

1. INTRODUCTION

Having clarified the thesis of revelation in chapter 4, in this chapter I turn to address its incompatibility with physicalism. Discussion in this chapter will help bring out the explanation, to be proposed in chapter 7, of the intuition of dualism in terms of the intuitiveness of revelation.

It is worth noting that in the literature, while revelation is often thought of as being incompatible with physicalism (Lewis 1995; Goff 2017), the alleged incompatibility is controversial (Damjanovic 2012; Trogdon 2017; Stoljar 2018). Of course, whether there is an incompatibility between revelation and physicalism, and how this incompatibility is to be understood provided that there is one, depends on how we explicate the thesis of revelation and the doctrine of physicalism. My formulation of the thesis of revelation in chapter 4 was as follows:

(Revelation)

By having an experience-token with phenomenal property Q, one is in a position to know a truth, namely, 'Q is X', where the predicate 'X' captures the essence of Q, although it may be hard to put into words.

In spelling out the incompatibility between revelation and physicalism, I shall focus on standard versions of physicalism ('standard physicalism' for short), which most physicalists would agree to. According to standard physicalism, phenomenal properties of conscious experiences are purely physical or functional – a physical/functional description exhausts the nature of qualia. So according to physicalism standardly conceived, the only truth that captures the essence of Q is some physical or functional truth, i.e. 'Q is X', where 'X' is a physical or functional predicate, as in, for instance, 'Phenomenal red is R-fibres firing in the brain', 'Itchiness is the property of having some physical property or other that plays the functional role F', etc. Standard physicalism thus takes qualia to have physical or functional essences. (A note on terminology: in Part I, I used the term 'physical' to include 'the functional'. The claim that consciousness is not physical thus entails that consciousness is not functional. In this chapter, I shall use the term 'physical' in a way that excludes 'the functional'. This helps to delineate different versions of physicalism more clearly.)

The structure of this chapter is as follows. In section 2, I address Lewis' remarks on the incompatibility between revelation and physicalism. This discussion clarifies how revelation is incompatible with the claim that qualia have *physical* essences. In section 3, I turn to consider role-functionalist versions of physicalism. I shall argue that revelation is incompatible with role functionalism and with the claim that qualia have *functional* essences. In section 4, I lay out a formulation of the incompatibility between the thesis of revelation and standard versions of physicalism. Section 5 concludes the chapter.

2. LEWIS ON REVELATION AND PHYSICALISM

Lewis (1995: 142) thinks that the thesis of revelation is part of the folk psychological concept of qualia and 'seems obvious' (that is, intuitive). However, he rejects revelation because it is incompatible with physicalism, which he commits to. After spelling out the thesis of revelation, Lewis moves on to discuss how physicalists cannot accept revelation.

He explains:

If, for instance, *Q* is essentially the physical property of being an event of C-firing, and if I identify the qualia of my experience in the appropriate 'demanding and literal' sense, I come to know that what is going on in me is an event of C-firing. Contrapositively: if I identify the quale of my experience in the appropriate sense, and yet know nothing of the firing of my neurons, then the quale of my experience cannot have been essentially the property of being an event of C-firing.

A materialist cannot accept the Identification Thesis. If qualia are physical properties of experiences, and experiences in turn are physical events, then it is certain that we seldom, if ever, identify the qualia of our experiences. Making discoveries in neurophysiology is not so easy! (Lewis 1995: 142)

The idea here should be straightforward. Recall that for Lewis (1995: 142), the term 'qualia' is 'a name for the occupants of a certain functional role that is spelled out in our tacitly known folk psychology'. We have seen that Lewis allows that the thesis of revelation (his 'identification thesis' – component (iv) in section 3 of chapter 4) describes one component of the qualia-role. Let *Q* be the phenomenal property of a pain-experience (its painfulness), and let *X* be *Q*'s essence and the predicate '*X*' capture *Q*'s essence. Physicalists like Lewis take *X* to be a physical property that plays the functional role of quale *Q*. So '*X*' is a physical predicate, e.g. 'being an event of C-fibres firing.' But if the functional roles of qualia include the thesis of revelation – that is, include our being

in a position to know the essences of qualia simply by being the subject of experiences that have the qualia among their properties – then we will not be able to find any physical properties that fulfil those functional roles. Corresponding to quale Q, there is no physical predicate 'X' that captures the essence of Q and is such that, just by having an experience with quale Q, one is in a position to know the truth 'Q is X'. In having a pain-experience, I am certainly not in a position to know *de dicto* that the painfulness of my experience of pain is its being an event of C-fibres firing. What this means is that no physical property can play the full qualia-role, because no physical property can play the revelation component of the qualia-role.

As a result, Lewis drops the thesis of revelation from the full-blown functional definition of qualia which we saw at the beginning of section 3 in chapter 4. What remains then are components (i), (ii) and (iii):

- (i) Qualia cause our 'abilities to recognize and to imagine experiences of the same type'.
- (ii) Qualia are 'responsible for responses of pleasure, disgust, etc.'.
- (iii) Qualia are 'responsible for judgements of similarity-distance, e.g., the judgement that these two colour-samples nearly match whereas those two contrast strikingly' (Lewis 1995: 141).

These remaining components constitute the nearest functional role which some physical properties can indeed satisfy. So, to the question ‘Should a materialist believe in qualia?’,

Lewis responds:

Yes: he should believe in imperfect but good-enough deservers of the name, occupants of the part of the folk-psychological role we get by leaving out the Identification Thesis [i.e. revelation]. And no: he should not believe in perfect deservers of the name, occupants of the entire role. (Lewis 1995: 142-3).

As a physicalist, Lewis denies that there is anything that *perfectly* satisfies our folk-psychological conception of qualia. But he proposes that physical properties of physical events (such as the property of being an event of C-fibres firing) can deserve – *imperfectly*, but well enough – the name ‘qualia’.

3. REVELATION AND FUNCTIONALISM

As we saw, Lewis is a functionalist who takes qualia to be occupants of functional roles spelt out in folk psychology. As we also saw, Lewis subscribes to what is known as ‘the identity theory’, according to which qualia are just physical properties such as being an event of C-fibres firing. Lewis is thus a realiser functionalist, who identifies qualia with physical properties that realise the corresponding functional roles (see also Lewis 1966, 1980).¹ *Realiser functionalism* is distinguished from *role functionalism*, which identifies

¹ Lewis is also an a priori physicalist, who holds the view that the conditional ‘If P then C,’ where P stands for the totality of physical truths and C stands for a given phenomenal truth, is true *a priori*. One can also be an a posteriori physicalist and hold the view that ‘If P then C’ is only true *a posteriori*. Many proponents of a posteriori physicalism explicitly reject revelation (see Loar 1997; McLaughlin 2001, 2003; Papineau 2007, forthcoming; Balog 2012). Like Lewis, they reject it simply because it renders physicalism false.

qualia with functional role properties that may be multiply realised by different physical properties playing the same functional role (e.g. in differently constituted creatures). In our discussion on Lewis, we saw that revelation is incompatible with realiser-functional physicalism and the claim that qualia have *physical* essences. More needs to be said about whether revelation is also incompatible with role-functionalist physicalism and the claim that qualia have *functional* essences.² In the remainder of this section, by ‘functionalism’, I simply mean role functionalism.

Note that different versions of functionalism understood the notion of a functional role differently. According to scientific functionalism or psychofunctionalism, in the terminology of Block (1978), functional roles are specified in the vocabulary of cognitive science or neuroscience. According to what is known as ‘commonsense functionalism’, functional roles are specified in the non-technical, everyday vocabulary of commonsense psychology. It is straightforward that revelation is incompatible with psychofunctionalism. In having an experience with a particular quale Q, there is no functional property X cashed out in technical, scientific vocabulary such that just by having the experience alone, one can know that Q is X. It is not so straightforward that the thesis of revelation is incompatible with commonsense functionalism. All else equal, being in a position to know a piece of everyday commonsense seems more likely than being in a position to know something technical and scientific.

² Both identity theorists and realiser functionalists would agree that phenomenal properties have physical essences. In contrast, role functionalists take phenomenal properties to have functional essences.

A commonsense functionalist would say that 'X' is a functional predicate couched in the non-technical vocabulary of commonsense psychology. She might also insist that one is indeed in a position to know such a functional truth 'Q is X' by having an experience with quale Q.³ However, I shall argue that revelation is incompatible with commonsense functionalism.⁴

In formulating and explicating the thesis of revelation (chapter 4, section 4), I mentioned the idea that the essence-capturing truth 'Q is X' may be hard to put into words. This is supposed to capture the intuitive idea that phenomenal experiences have indescribable phenomenology.⁵ Now commonsense functionalists claim that qualia can be fully defined by their functional roles. One might think that intuitively, such functional roles do not seem to capture the 'je ne sais quoi (I know not what)' character of qualia. But if the essence of a quale is functional (or physical), then it can be put into words. Commonsense functionalists would of course deny that any qualia are indescribable and insist that if one states the full functional role of a quale in the everyday vocabulary of commonsense psychology, then one has successfully given a definition of what that quale is essentially. In spelling out the incompatibility between revelation and commonsense functionalism, I shall grant that the thesis of revelation does not entail, but merely allows, that the essence-capturing truth 'Q is X' is hard to put

³ I am indebted to Frank Jackson for pressing me on this issue.

⁴ There are problems with commonsense functionalism that are independent of the thesis of revelation (see Block 1978).

⁵ The idea that qualia are indescribable can be traced to Descartes, who thinks that sensations have an 'I-know-not-what' (*'je ne sais quelle'* in French) aspect (see Cottingham 2000). For a recent discussion on the indescribable character of experience, see Camp 2006.

into words.⁶ I shall also set aside the general concern that, intuitively, a state's having a certain functional role does not guarantee that the state have any phenomenal character at all.⁷

In order to satisfy the thesis of revelation, the commonsense functionalist would have to say that *by having a pain experience with phenomenal property Q* (where Q is the painfulness of pain), a subject is in a position to know the commonsense functional role of Q. That is, the subject is in a position to achieve *de dicto* knowledge of the form 'Q is X', where 'X' specifies the commonsense functional role of painfulness, which captures the essence of painfulness. I shall argue that *in having a pain experience with phenomenal property Q*, one is not in a position to know the commonsense functional role of Q; one is not in a position to know *de dicto* such a truth 'Q is X', where the predicate 'X' captures the functional essence of Q in commonsense psychological terms.

Commonsense functionalists define qualia in terms of their functional roles by drawing on what is common knowledge about mental states with these qualia (Braddon-Mitchell and Jackson 2007: 52). According to these functionalists, this common knowledge has three parts: 'there are input clauses – clauses that say what sorts of events cause mental states in people; output clauses – clauses that say what sort of behaviours

⁶ The idea that mere words are unable to communicate the essence of phenomenal property Q is *not* officially part of the thesis of revelation. In explicating revelation, I merely left open the possibility that the essence-capturing truth 'Q is X' *may* be hard to put into words. As we have seen, the reason that the identity theory or realiser functionalist physicalism is incompatible with the thesis of revelation is not that revelation requires the essences of qualia to be indescribable – it does not explicitly require that.

⁷ This worry was raised in chapter 2 in discussion of the a priori physicalists' proposal for bridging the explanatory gap, which relies on a functional analysis of qualia.

are caused by mental states; and internal role clauses – clauses that describe the internal interactions of mental states’ (Braddon-Mitchell and Jackson 2007: 48). The input clauses for quale pain would include the clause ‘bodily damage causes experiences with quale pain’; output clauses would include the clause ‘experiences with quale pain cause bodily movements that relieve pain and minimize damage’; internal clauses would plausibly include the clause ‘experiences with quale pain cause desires to relieve pain, which typically lead to desires to take painkillers (if one also believes that taking painkillers would be likely to relieve the pain)’ (Braddon-Mitchell and Jackson 2007: 52-3). Giving a definition of what quale pain is in commonsense functional terms requires specifying all three kinds of clauses.

As we saw in section 4 of chapter 4, according to the thesis of revelation, the knowledge *de dicto* that Q is X, where the predicate ‘X’ captures the essence of Q, is gained *in virtue of* having an experience with Q. Let us agree that one is in a position to know the tripartite commonsense functional role of quale pain. Still, it seems implausible that it is *in virtue of having a pain experience*, that one is in a position to have such knowledge, which, according to commonsense functionalists, captures the essence of quale pain. Consider, first, the internal clauses which say that the painfulness of pain causes a desire to relieve pain and motivation toward actions that the subject believes would make the painfulness of pain more likely to stop. The details of such actions and motivations (to take painkillers, for example), are unlikely to be known just in virtue of having a painful experience. If ‘causing the desire to take painkillers’ is built into the internal clauses that specify part of the functional role of quale pain, then the functional

role will not be *revelation compatible*, as we shall say. It is just not true that in virtue of having a pain experience alone, one is in a position to know that quale pain is X where the predicate 'X' includes 'painkillers'. Of course, the commonsense functionalist has a ready response to this point. They should say that what is core and primary in the internal clauses of the functional role of quale pain is nothing specifically about painkillers, but just that the painfulness of pain causes the desire that it stop.

Consider, second, the input clauses of the functional role of quale pain which say that bodily damage causes experiences with quale pain. It does not seem to be the case that in virtue of having a pain experience, that is, in virtue of feeling pain, I am in a position to know that it is bodily damage that causes my experience to have a particular phenomenal character. Nothing about bodily damage is revealed to one in having pain experiences like headaches, stomachaches, cramps etc. Even in cases where one feels pain in one's finger as a result of its being cut, having the sensation of pain in one's finger does not on its own reveal that one has a cut on one's finger which gives rise to the sensation. According to proponents of commonsense functionalism, the input clauses that form part of the definition of quale pain explicitly mention 'bodily damage' (Braddon-Mitchell and Jackson 2007: 52). But a functional role with such input clauses will not be revelation compatible. In having the phenomenal experience of pain *alone*, one is not typically in a position to know that quale pain is X where the predicate 'X' includes 'bodily damage'.

Now, we can grant that there are commonsense clauses about quale pain that one is in a position to know in virtue of having a painful experience, including the output

clause 'the experience with quale pain causes one to wince, to have reflexive movements of avoidance or withdrawal', the internal clause 'the experience with quale pain causes one to want the experience to stop', and perhaps others. But it is far from obvious that the clauses of this reduced, revelation-compatible functional role specification are adequate to capture the *essence* of quale pain. Specifically, these functional role clauses might not distinguish the painfulness of pain from other qualia. For example, the phenomenal properties of experiences such as biting into a slice of lemon, seeing something extremely bright, hearing a loud fire alarm going off, etc. all typically cause one to wince and to have reflexive movements of avoidance or withdrawal. In all these cases, the subject also typically desires the relevant experience to stop. But while these experiences are examples of distress, they are typically not considered as experiences of pain – one would not usually say that one is in pain if one has one of these experiences.

The claim that commonsense role functionalism is compatible with the thesis of revelation clearly faces a dilemma. If a functional role specification is to capture the essence of quale pain, then it must individuate quale pain, in the sense that no other quale has that functional role. A functional role specification that includes enough of what is commonly known about the role of pain to meet this individuation requirement is plausibly not revelation compatible. And a functional role specification that is sufficiently reduced to be revelation compatible is plausibly no longer individuating. This pattern certainly generalizes beyond the example of the painfulness of pain.

Consider how a commonsense functionalist would spell out the functional roles of the phenomenal properties of various color experiences. In his paper 'Naming the

Colours', Lewis (1997: 327) mentions the following two clauses as being central to the commonsense functional roles of worldly colors and corresponding properties of color experiences:

D1 *Red* is the surface property of things which typically causes
experience of red in people who have such things before their eyes.

D2 *Experience of red* is the inner state of people which is the typical effect
of having red things before the eyes.

D1 and *D2* inter-define *red* and *experience of red*. But exactly similar clauses inter-define *green* and *experience of green*. So the putative functional role of the pair $\langle \text{red}, \text{experience of red} \rangle$ does not individuate that pair. The pair $\langle \text{red}, \text{experience of red} \rangle$ is very far from being the only pair $\langle G, H \rangle$ such that *G* is the surface property of things which typically causes *H* in people who have such things before their eyes and *H* is the inner state of people which is the typical effect of having *G* things before the eyes.

In order to break out of the circle of inter-definition and distinguish experience of red from other colour experiences, *D2*, which is the input clause for experience of red, needs to be augmented with something about common examples of worldly surfaces that are red, and so (according to *D1*) cause experience of red. But it seems that having an experience of red does not, by itself, put the subject in a position to know about such examples. For instance, one might define *experience of red* as being 'typically caused by the color of a British pillar box' (Lewis 1997: 335). But surely, in having an experience of red alone, one is not in a position to know anything about British pillar boxes nor in a position to know that phenomenal red is typically caused by seeing a British pillar box

This second example helps to sharpen the dilemma facing commonsense functionalists who claim that their position is compatible with the thesis of revelation. If commonsense functionalists define qualia in functional terms by appealing to everyday concepts of worldly objects and states like 'painkillers' or 'bodily damage' in the case of quale pain and 'British pillar box' in the case of phenomenal red, then commonsense functionalism is incompatible with revelation. But if commonsense functionalists avoid making references to these things in spelling out the commonsense functional roles of qualia, then it seems no longer clear that, with only the vocabulary that would *not* create a problem for maintaining revelation, they can give adequate functional definitions that individuate qualia. It does not seem that the reduced functional definition of a given quale Q, which the subject is in a position to know by having an experience with quale Q, is fine-grained enough to distinguish Q from other qualia.

Given this dilemma, our conclusion must be that commonsense role functionalism – like realiser functionalism (and the identity theory) and scientific role functionalism (psychofunctionalism) – is incompatible with the thesis of revelation.

4. THE REVELATION ARGUMENT AGAINST PHYSICALISM

At this point, it would be useful to formally lay out the incompatibility between the thesis of revelation and physicalism (in its identity theory and realiser and role functionalist versions). Here I will present the incompatibility between the two in the format of an argument, namely, an argument from the thesis of revelation to the falsity of physicalism. Let's call this argument 'the revelation argument'. (Again, physicalism

is here understood as ‘standard physicalism’, which is the claim that qualia have physical or functional essences.⁸ If there are positions that do not entail this claim, but are nonetheless worthy of the name ‘physicalism’, then whether those versions of physicalism are incompatible with revelation is a separate question.⁹)

- (1) If S has an experience-token with quale Q, then S is in a position to know *de dicto* that ‘Q is X’, where the predicate ‘X’ captures the essence of Q.
- (2) If physicalism is true, then qualia have physical/functional essences.
- (3) If qualia have physical/functional essences, then by having an experience-token with quale Q, S should be in a position to know *de dicto* that ‘Q is X’, where ‘X’ is a physical/functional predicate which captures the essence of Q.
- (4) It is not true that by having an experience-token with quale Q, S is in a position to know *de dicto* that ‘Q is X’, where ‘X’ is a physical/functional predicate which captures the essence of Q.

⁸ Here the word ‘physical’ is still used in the sense that excludes ‘the functional’.

⁹ For instance, one might argue that physicalism defined as a grounding thesis does not entail that qualia have physical/functional essences (see Stoljar 2018). Whether revelation is incompatible with grounding physicalism, i.e. physicalism cashed out as a grounding thesis, is an issue I will discuss in chapter 8. In this chapter, I am not concerned with whether or not revelation is incompatible with all versions of physicalism. My goal is to spell out how revelation is incompatible with the claim that phenomenal properties are physical/functional, which is what standard versions of physicalism subscribe to. This incompatibility will help us explain why we have the intuition of dualism, the intuition that consciousness is nonphysical/nonfunctional.

(5) Physicalism is false.

Premise (1) is entailed by the thesis of revelation (see the second paragraph of this chapter). (2) states what the essences of qualia would be if physicalism were true. (3) follows from (1) and (2). (4) is supported by the considerations discussed in sections 2 and 3. (5) then concludes from (3) and (4) that physicalism is false.

Given our clarification of the distinction between knowledge *de dicto* and knowledge *de re* of qualia in section 5 of chapter 4, it is worth noting that physicalists, in denying revelation, can still say that by having an experience with quale Q, one is in a position to know *de re of* X, where the predicate 'X' captures the essence of Q, that Q is it (see Lewis 1995: 143). A physicalist might say that there are two distinct modes of presentation of the essence of the *phenomenal* property Q – a *physical/functional* mode of presentation and a *phenomenal* mode of presentation, and that the mode of presentation that reveals the essence of Q is the physical/functional one (see Loar 1997: 603). So 'X', which captures the essence of Q, is a physical/functional predicate. Such a physicalist can say that by having an experience with quale Q, one is in a position to know *de re of* physical/functional property X (but under its phenomenal mode of presentation) that Q is it, e.g. know *de re* of the property of being an event of C-fibres firing that the painfulness of a pain experience is that property.¹⁰ This is analogous to the case of Tom's knowledge of Venus (chapter 4, section 5). Suppose again that Tom thinks of Venus *as*

¹⁰ A physicalist can say that the essence of Q, which is a physical/functional property, can be thought about in a physical/functional way, i.e. *as* X, or in a phenomenal way, *as*, say, Y. Such a physicalist can say that by having an experience with quale Q, one is in a position to know *de dicto* that Q is Y. Note that 'Y,' which is a phenomenal predicate, refers to but (according to this physicalist account) does not *capture* the essence of Q.

Phosphorus and not *as Venus* and that Tom knows *de dicto* that Hesperus is Phosphorus. In such a case, we can still say that ‘Tom knows *of Venus* that Hesperus is it’. Tom has knowledge *de re of Venus* (even without any idea what Venus is) insofar as he thinks of Venus *as Phosphorus*.

One can also see that the doctrine of the simplicity of qualia (chapter 4, section 6) is in tension with physicalism. If qualia are physical or functional properties, which are presumably not simple, then qualia cannot be simple. Hence, physicalists also ought to reject the idea that qualia are simple (see Lewis 1995: 142). In section 6 of chapter 4, we saw that there is an argument from revelation to the conclusion that some qualia are simple. The argument crucially relies on the premise that there is at least a range of cases such that in having an experience with Q, one is not in a position to know that Q has internal structure. A commonsense functionalist who claims their position to be compatible with revelation is likely to reject this premise and claim that in all cases, by having an experience with Q, one is indeed in a position to know the complex structure of Q as a result of being in a position to know the complex commonsense functional role of Q. This allows such a commonsense functionalist to reject the entailment from the thesis of revelation to the claim that some qualia are simple. But as we argued in the previous section, commonsense role functionalism is incompatible with the thesis of revelation in any case.

It is worth noting that arguments against physicalism mainly come in two forms – the knowledge argument (e.g. Jackson 1982, 1986) and the conceivability argument (e.g. Chalmers 1996). Here we also see a direct argument from the thesis of revelation against

physicalism. Despite Lewis' illuminating discussion in his 1995 paper, it is only more recently that the thesis of revelation has been brought up explicitly in relation to the discussion of whether physicalism is true (Goff 2017; Trogdon 2017; Majeed 2017; Stoljar 2006, 2009).^{11 12} But in the literature, revelation and physicalism are not always thought

¹¹ One does find something similar to the thesis of revelation being used as an objection against the phenomenal concept strategy (see Horgan and Tienson 2001; Levine, 2001: 84, 2007; Nida-Rümelin, 2007; Goff, 2011, 2015; Veillet 2015). This is the worry that the phenomenal concept strategy cannot account for the idea that phenomenal concepts afford us substantive knowledge about the nature of phenomenal properties. For replies to this challenge from a posteriori physicalists, see Schroer (2010), Díaz-León (2014), Elpidorou (2016), and Taylor (2017).

¹² Apart from Lewis, two other philosophers are worth considering when situating the thesis of revelation as a consideration against physicalism in the history of analytic philosophy: Kripke and Nagel.

In *Naming and Necessity* (1980), Kripke argues against physicalism. Lewis (1995) interprets Kripke as relying on the thesis of revelation, but Stoljar (2009) disagrees. I am inclined to agree with Stoljar here. I understand Kripke's argument as crucially relying on two points: first, the intuition that pain and C-fibre stimulation seem to be contingently related; and second, that such an intuition cannot be explained away. Kripke (1980: 155) then draws the conclusion that physicalism is problematic. Nothing in his argument explicitly relies on the thesis of revelation. However, revelation might nevertheless play a role in Kripke's conceiving the counterintuitiveness of physicalism. Had he thought, like Lewis, that revelation was false despite being intuitive, he would have realised, I think, that a physicalist could appeal to the intuitiveness of revelation to explain away the intuition of contingency regarding the experience of pain and C-fibre stimulation.

In his 1974 paper, Nagel questions the idea that experiences can have an objective nature. He (1974: 448) writes:

Very little work has been done on the basic question (from which mention of the brain can be entirely omitted) whether any sense can be made of experiences' having an objective character at all. *Does it make sense, in other words, to ask what my experiences are really like, as opposed to how they appear to me?* (italics added)

Nagel doubts that there is such an appearance-reality distinction with respect to experience, that is, that there is an objective reality of experience beyond its appearance or subjective character. He (1974: 444-5) writes:

The idea of moving from appearance to reality seems to make no sense here. ... Certainly it *appears* unlikely that we will get closer to the real nature of human experience by leaving behind the particularity of our human point of view and striving for a description in terms accessible to beings that could not imagine what it was like to be us. *If the subjective character of experience is fully comprehensible only from one point of view, then any shift to greater objectivity – that is, less attachment to a specific*

to be incompatible (e.g. Damjanovic 2012; Trogdon 2017; Majeed 2017; Stoljar 2018).¹³ Of course, the issue here depends on how one explicates the thesis of revelation and the doctrine of physicalism. We have just seen that the thesis of revelation, as we have formulated it in chapter 4, is indeed incompatible with standard versions of physicalism, which take qualia to have physical/functional essences. Whether or not the revelation argument is convincing depends on whether the thesis of revelation is true. I will address objections to the thesis of revelation and relate it to the metaphysical debate on the nature of consciousness in the final two chapters of the dissertation.

5. CHAPTER CONCLUSION

In this chapter, we spelt out the incompatibility between revelation and physicalism. Revelation is incompatible with the claim that qualia have physical or functional essences. The latter is precisely what standard versions of physicalism subscribe to.

viewpoint – does not take us nearer to the real nature of the phenomenon: it takes us farther away from it. (second italics added)

Nagel seems to say that if we *fully comprehend* the subjective character of our experience from our subjective point of view, that is, if all there is to know about (the nature of) phenomenal properties is known from the subjective point of view, then phenomenal properties themselves do not have an objective nature capable of being given an objective description. So, experiences, that is, conscious experiences-types, do not have an objective nature. One might read Nagel as appealing to the thesis of revelation in the italicised sentence. I think this reading is reasonable. In questioning the idea that experience has an objective nature, Nagel seems to be relying on the intuitiveness of revelation, the thought that we seem to already know all there is to know about the nature of the subjective character of experience from the subjective point of view.

¹³Those who doubt the incompatibility of revelation and physicalism either understand the thesis of revelation somewhat differently from the one under discussion (e.g. Damjanovic 2012; Trogdon 2017), or would question whether the thesis of revelation as I have formulated here rules out all versions of physicalism (e.g. Stoljar 2018).

As we also saw in this chapter, Lewis thinks that revelation is an intuitive claim and rejects it only because it is incompatible with the version of physicalism he subscribes to. The claim that revelation is intuitive is in fact controversial (Stoljar 2009). In the next chapter, I will put forward a linguistic argument for the intuitiveness of revelation. The conclusion of the next chapter, together with the conclusion established in this chapter, namely, that revelation is incompatible with the claim that qualia have physical/functional essences, helps to bring forth the explanation of the intuition of dualism to be proposed in chapter 7.

CHAPTER 6

THE INTUITIVENESS OF REVELATION: A LINGUISTIC ARGUMENT

1. INTRODUCTION

In the last chapter, we saw that the thesis of revelation is incompatible with standard versions of physicalism, which take phenomenal properties to have physical or functional essences. The thesis of revelation is often thought of as *intuitive*, or in Chalmers' words, as having 'a certain intuitive plausibility' (2016: 190). In claiming that the thesis of revelation is intuitive, philosophers usually mean that it *seems* true; they do not mean that all things considered, we should think that it *is* true. Many physicalists, who endorse the incompatibility of revelation and physicalism, also think that revelation is intuitive. Lewis (1995: 142) thinks that revelation (i.e. what he calls 'the identification thesis') 'seems obvious'. Papineau (forthcoming) thinks that revelation 'is a highly intuitive idea'. According to McLaughlin (2003: 99), '[t]he powerful intuitive appeal of the doctrine of Revelation for what it's like to see colours—for the phenomenal characters of colour experiences—seems ... undeniable'. According to Hill (2014: 199-200), 'we are inclined to think that experiential awareness provides us with full access to the essential nature of qualia. Our grasp of them is not perspectival or limited in any way'. Braddon-Mitchell (2007: 287) thinks that because revelation is so intuitive and is incompatible with physicalism, 'there is a big intuitive cost in accepting physicalism'. All these philosophers are avowed physicalists and acknowledge the incompatibility between revelation and their physicalist stance, yet they all take revelation to be intuitive.

Of course, a claim may be said to be intuitive in different senses. Sometimes, a particular claim is said to be intuitive because it seems true on simple pre-theoretical reflection.¹ For instance, it is intuitive that a heavy spherical object would make a concave shape in soft sand, and the intuitiveness of this claim is evident upon simple reflection. But not all claims that can be said to be intuitive seem true upon only simple reflection. For instance, a complex mathematical proof may be said to be intuitive, in the sense that those who understand the proof can see clearly how the proof works and how it successfully proves what it sets out to prove. A claim might also be said to be intuitive because it is part of the ordinary conception of a relevant domain. If a claim P is part of the ordinary conception of some domain, then P is a claim that an ordinary, sensible person would assent to, or be inclined to assent to. For instance, it is intuitive that there are medium-sized objects like tables and chairs – ordinary people would indeed assent to the claim that there are such medium-sized objects.

One might think that the thesis of revelation seems true upon simple reflection about our experience. Although I am sympathetic to the latter claim, I am not sure if everyone would agree. The thesis of revelation certainly doesn't seem intuitive in the sense that a mathematical proof can be thought to be intuitive. Here, in discussing the

¹ To pre-theoretically reflect on something is to think about it without bringing in theories or theoretical considerations that may be relevant to thinking about the phenomenon itself. Although it is hard to precisely draw the distinction between pre-theoretical thinking and theorising, and indeed perhaps there is no clear-cut distinction between the two, the distinction is nevertheless intuitive. For instance, it is intuitive that the world we experience is mind-independent and external to us, and this much is evident upon our pre-theoretical reflections about our experience of the world. But if one starts entertaining sceptical scenarios, it is no longer clear that the belief that the world we experience is mind-independent from us is in fact true.

intuitiveness of revelation, I will focus on the third sense of intuitiveness, i.e. being part of the ordinary conception of the relevant domain.² Indeed, among those who take the thesis of revelation to be intuitive, many explicitly associate its intuitiveness with its being part of the ordinary conception of experience. For instance, according to Lewis, revelation ‘seems obvious because it is built into folk psychology’ (1995: 142).³ For Braddon-Mitchell (2007: 287), revelation is intuitive because ‘[i]t is a fairly deep feature of our conception of consciousness (or at least of qualia)’.

However, the issue of whether or not the thesis of revelation is part of the ordinary conception of experience is contentious. Stoljar (2006: 125; see also 2009) has argued extensively that there is no good reason to think that ‘the conception of experience that incorporates revelation is the ordinary one’. Among defenders of physicalism against *the revelation argument* who think that the thesis of revelation ought to be rejected, there is then a difference between those who do not think that revelation is part of our ordinary conception of experience and those who do. The difference is

² Throughout this chapter, I take it for granted, as it is usually assumed in the literature, that the notion of *phenomenal property* or *phenomenal type* is a folk psychological concept, or at least a concept that ordinary people can easily grasp. The question under consideration is whether the thesis of revelation is part of the folk or ordinary conception of experience and its phenomenal properties.

³ One might question whether whatever is part of *folk psychology* would seem obvious as Lewis claims (Stoljar 2009). The issue here depends on how folk psychology is conceived, whether as a set of platitudes, i.e. *the platitude account* (Lewis 1972: 256), or as abstract rules and generalisations, i.e. *the theory account* (Lewis 1994: 416). On the theory account, what is part of folk psychology does not necessarily seem obvious, e.g. grammatical rules which involve technical concepts (Stoljar 2009). It is possible that in writing the paper ‘Should a Materialist Believe in Qualia?’, Lewis still held the platitude account of folk psychology, though he later changed his mind and adopted the theory account. Here, I shall avoid the term ‘folk psychology’, and only speak of what belongs to the ‘ordinary conception of experience’.

reflected in the options they adopt regarding their rejection of revelation (see Stoljar 2009):

[Option 1]

Reject revelation by rejecting part of our ordinary conception of experience:

the thesis of revelation is part of the ordinary conception of experience, but physicalists need to reject that part of the ordinary conception of experience.

[Option 2]

Reject revelation outright: there is no need to take revelation seriously because it is not even part of the ordinary conception of experience.

Those physicalists who take revelation to be incorporated in our ordinary conception of experience adopt Option 1.⁴ They would say that there is no physical property that *perfectly* satisfies our ordinary conception of experience, but physical properties like being an event of C-fibres firing *imperfectly* satisfy our ordinary conception of experience – they perfectly satisfy a replacement conception of experience that does not incorporate the thesis of revelation. As we have already seen, this is the line of thought one finds in Lewis (1995: 142-3). For these physicalists, there is a cost that comes with denying revelation insofar as failing to accommodate certain ordinary intuitions is considered to be undesirable for a philosophical theory (Braddon-Mitchell 2007: 287). Option 2 is adopted by those physicalists who think that the thesis of revelation is not even part of

⁴ Option 1 is adopted by physicalists, e.g. Lewis, who are associated with what is known as ‘the Canberra Plan’.

the ordinary conception of experience and need not be taken seriously. For them, denying revelation does not even come with a cost.

As already mentioned, Stoljar, *contra* Lewis, does not think that revelation is part of our ordinary conception of experience. Stoljar (2009) considers a number of potential reasons for which one might think that revelation is part of the ordinary conception, and argues that none of these reasons are actually persuasive. I will not rehearse Stoljar's arguments here. Instead, I will put forward a new consideration in support of the claim that the thesis of revelation is in fact part of the ordinary conception of experience. I shall argue that this consideration is persuasive, and hence revelation is also intuitive for this reason.

But before delving into my argument, let me first clarify a preliminary issue. One might think that revelation cannot possibly be part of our ordinary conception of experience because it features the notion of *essence*, which some might think is technical and not pre-theoretically clear (e.g. Stoljar 2009: 127).

I think this worry is unwarranted. The notion of essence which we have been using here, namely, 'that which makes something the thing it is', is the Aristotelian/Lockean/Finean *real definitional* account of essence. Fine takes the notion of essence to be akin to that of definition, and to be such that we can all easily get a grip on it by considering examples. Of course, this does not mean that we all know clearly what the essence of any given thing is. There is a distinction to be drawn between having an adequate understanding of the notion of essence itself, and being able to answer the question – *What* makes a given thing the thing it is? One might very well have the former

understanding without being able to answer the latter question. To determine what counts as the essence of a thing is often an a posteriori enquiry. But this does not mean that the folk lack an adequate understanding of the notion of *essence* itself.

Furthermore, studies in developmental psychology have shown that humans from a young age have the tendency to represent things and categories of things to have hidden essences (e.g. Medin and Ortony 1989; Keil 1989; Gelman 2003; Bloom 2010; Newman and Knobe forthcoming). This is known as ‘psychological essentialism’. The concept of essence that these psychologists invoke is precisely the Aristotelian/Lockean/Finean notion of *real essence* (see Gelman 2003: 3). Psychological essentialism thus suggests that the concept of essence is a concept that the folk have and frequently utilise.

So, while the English word ‘essence’, understood in the sense at issue here, might seem to be a technical term, the concept is nevertheless intuitive and part of folk psychology.⁵ Having clarified this crucial point regarding the folk-theoretical status of the notion of essence, let’s now turn to my argument in support of the claim that the thesis of revelation is part of our ordinary conception of experience.

2. THE ARGUMENT

A good starting point for investigating whether revelation is part of our ordinary conception of experience is language, in particular our ordinary discourse. If revelation

⁵ It is interesting to note that in Chinese, the word for ‘essence’ in the sense under consideration here, i.e. *ben zhi*, is common parlance.

is part of our ordinary conception of experience, then it would probably be evident in the way we talk about experience. In the rest of the chapter, I present some linguistic data, and argue that the best explanation of the data is that the thesis of revelation is part of our ordinary conception of experience.

Consider the following sentences:

- (1) I know what an emerald looks like, but I don't know what an emerald *really* is.
- (2) I know what an itch feels like, but I don't know what an itch *really* is.
- (3) I know what an itch feels like, but I don't know what the feeling of an itch *really* is.

It seems that (1) sounds perfectly fine. I suspect that to an ordinary English speaker (2) sounds quite odd, and (3) is likely to sound even odder than (2).⁶

Consider also the following set of sentences.

⁶ Note that without the word 'really' in each of the second conjuncts, one might think that all three sentences sound odd, including (1). If 'really' is elided in (1), the resulting sentence would raise the question: How did you know it was an emerald that looks like thus-and-so, if you don't know what an emerald is? The thought here is that if one knows what an emerald looks like, then one would naturally know what an emerald is in the everyday, undemanding sense of 'knowing what something is'. The sentence – 'I know what an emerald looks like, but I don't know what an emerald is' – would not sound odd given a somewhat different interpretation, e.g. if the interpreter takes the second conjunct to indicate some more demanding kind of knowledge that goes beyond the thin, undemanding sense of knowing what something is. Here we are only concerned with sentences where the adverb 'really' is present and is not interpreted as a filler that can be elided.

- (4) You have all seen rubies, but do you know what a ruby *really* is?
- (5) You have all experienced toothaches, but do you know what a toothache *really* is?
- (6) You have all experienced toothaches, but do you know what the feeling of a toothache *really* is?

Like the first triplet, I suspect that (4) sounds perfectly fine, (5) sounds somewhat odd, and (6) sounds very odd.⁷

As we saw, (3) and (6) sound very odd but not (1) and (4), and not (2) and (5) to the same extent. In the following, I shall argue that our linguistic intuitions elicited by these sentences, in particular by (3) and (6), demonstrate that the thesis of revelation is part of our ordinary conception of experience. For simplicity, I will only focus on the first set of sentences, i.e. (1)-(3).⁸ The argument takes the following explicit steps. I first argue that there is a particular kind of reading or interpretation of (3), which I call ‘the

⁷ Here I am following the orthodox idea in linguistics that competent native speakers of a language should be equal authorities in judging whether particular sentence tokens are odd-sounding. What this means is that we need not conduct wide-ranging empirical surveys to test whether certain sentences are odd-sounding, although this might add persuasiveness to the point being made.

⁸ (1) and (2) form something close to a minimal pair – a pair of sentences that differ in only one respect. They have the syntactic structure: I know what X looks/feels like, but I don’t know what X really is. In (1), X is *an emerald*. In (2), X is *an itch*. (2) and (3) form a minimal pair. They both have the structure: I know what an itch feels like, but I don’t know what X really is (where X is either *an itch* or *the feeling of an itch*). Similar things can be said about the second group of sentences, i.e. (4)-(6). One could equally focus on the second group of sentences to argue, as I shall, that revelation is part of our ordinary conception of experience.

essential reading', and this reading allows (3) to be rendered odd.⁹ I then argue that this latter fact is best explained by the hypothesis that revelation is part of the ordinary conception of experience. By inference to the best explanation, I conclude that the thesis of revelation is indeed part of the ordinary conception of experience. As a result, we also have a reason to believe in its intuitiveness. The argument is summarised as follows:

(Premise I) The essential reading of (3) allows the sentence to be rendered odd.

(Premise II) The fact that the essential reading of (3) allows the sentence to be rendered odd is best explained by appealing to the hypothesis that the thesis of revelation is part of our ordinary conception of experience.

(Conclusion) The thesis of revelation is part of our ordinary conception of experience.

In the next section, I give a detailed argument for Premise I. I argue for Premise II in section 4. Given the structure of the argument, one can problematise the conclusion of the above argument by raising objections against either Premise I or Premise II. Section 5 addresses and replies to potential objections against the argument.

⁹ In saying that there is a particular *reading* of sentence (3), I am claiming that it is possible to *interpret* or understand (3) in a particular way.

3. PREMISE I

My argument for Premise I, which I shall spell out in detail in this section, takes the following form:

- (i) There is some reading of 'know what X really is' which allows (3) to be rendered odd.
- (ii) There are two kinds of reading of 'know what X really is': (a) the essential reading; (b) non-essential readings.
- (iii) A non-essential reading of 'know what X really is' in (3) does not render the sentence odd.
- (Premise I) The essential reading of (3) allows the sentence to be rendered odd.

Let me first say something to support premise (i). We have seen that (3) is odd. (3) is odd because there is a tension between the two conjuncts, in this case, between claiming to know what an itch feels like and then claiming not to know what the feeling of an itch *really* is. Note also that sentences (1) to (3) above all include the construction 'know what X really is'. It seems that if we want to know what explains our intuitions regarding these sentences, it is important to pin down what this construction – 'know what X really is' – means in English. One might plausibly think that it is our interpreting the phrase 'know what X really is', and also 'don't know what X really is', in a particular way that allows (3) to be rendered odd (where X = the feeling of an itch), but not (1) (where X = an emerald).

Given the aforesaid, we can put forward premise (i). Premise (ii) will be made clear in subsection 3.2 below. Premise (iii) will be argued for with examples in subsection 3.3. In subsection 3.1, I shall discuss various constructions we use in English to talk about things' having hidden underlying essences.

3.1. X IS (ESSENTIALLY/REALLY/ACTUALLY/IN FACT) Y

In ordinary English, there are a number of ways to talk about what the hidden underlying essence of something is, where this something is an ordinary phenomenon that we are familiar with. For instance, one can directly use the adverb 'essentially':

- (7) Lightning is *essentially* a sudden electrostatic discharge that typically occurs during a thunderstorm.

We can also use other adverbs or adverbial phrases:

- (8) Water is *in fact* molecules made up of two hydrogen atoms and one oxygen.
- (9) Colours are *actually/really* reflectance properties of the surfaces of objects.

Adverbial phrases are not necessary for making claims about something's essence. Sometimes we can just use the copula 'is', as in:

- (10) Heat *is* mean kinetic energy.

However, the above constructions can also be used in the same syntactic environment, i.e. noun phrase + copula (+ adverb) + noun phrase, to convey meanings

that are not semantically related to the notion of essence. For instance, the copula 'is' can be used simply to indicate that a particular thing is of a particular kind, as in:

- (11) Diamonds *are* a luxury good.

Being a luxury good is not a necessary or essential characteristic of diamonds.

The construction 'X is essentially Y' can also be used in a way such that Y is not part of the essence of X:

- (12) The banking industry is *essentially* a man's territory.

Here 'essentially' is used in much the same way as the word 'basically', to indicate an important feature of the thing at issue, where such a feature need not be an essential or even a necessary one. Being a man's territory is merely a contingent feature of the banking industry, but it may be a significant one. Because in English 'essentially' is commonly used synonymously with 'basically', 'primarily', etc., whereas 'really'/'actually'/'in fact' are not, replacing 'essentially' in sentences like (12) with one of these other adverbs would alter the meaning of the sentence.

Adverbial phrases like 'really', 'actually' and 'in fact' can also be used in ways that are independent from essence talk:

- (13) These marbles are *really/actually/in fact* worthless things.

'X is really/actually/in fact Y' means that X has the property of being Y and this property is contextually salient in some way, e.g. a property that is not expected or believed to be a property of X.

The upshot of this subsection is that all the locutions that we use to talk about hidden essences, i.e. 'X is (essentially/really/actually/in fact) Y', can also be used in ways that are not related to essence talk. These locutions can be used to mean simply that something has a property which, though not essential, is contextually salient in some way.

3.2. KNOW WHAT X REALLY IS

Given these different ways of expressing the essence of something, we can say that someone knows the essence of something with the following syntactic construction:

- (14) S knows what X (essentially/really/actually/in fact) is.

For ease of presentation, I will focus only on the following construction:¹⁰

- (15) S knows what X (really) is.

Let's first look at the simple construction 'S knows what X is'. This construction is often used without meaning that S knows what X's essence is, as in:

- (16) I know what jade is. I have seen it in the shops.

¹⁰ As mentioned, 'essentially' is frequently used in ordinary English to mean 'basically', which is not related to the notion of essence in philosophy. In ordinary English, 'in fact' occurring before 'is' in the construction 'S knows what X (adverb) is' might sound rather clunky. The word 'really' has the same meaning as 'actually' in the same syntactic environment, i.e. 'S knows what X (adverb) is'. 'Really' also has additional meanings. In such cases, the word occurs in different positions. For instance, 'really' can be used as an intensifier to mean 'very' as in 'The lake is *really* green' or 'She has *really* told a moving story'.

Knowing what something is does not in general require one to know the essence of that thing. If one has an adequate grasp of the notion of jade, or is competent to think about jade as such, or knows how to identify jade in normal circumstances (e.g. in jewellery shops), then one, in a sense, knows what jade is. Knowing what something is in this sense is like knowing who someone is:

(17) I know who Jimmy is.

Now, let's turn to the phrase 'S knows what X really is'. It can mean that S knows the underlying essence of X, as in:

(18) Every 10-year-old knows what water is, but not every 10-year-old knows what water *really* is. Claire, who had memorised the entire periodic table by age 8, knows what water *really* is.

Alternatively, 'S knows what X really is' may be used in ways that have nothing to do with the essence of X, as in:

(19) Gloria knows what these marbles *really* are: they are worthless things.

This use of 'S knows what something really is' is similar to 'S knows who someone really is' as in:

(20) Tom has become a regular guest at family gatherings, but only Claire knows who Tom *really* is: Tom is a con-man.

In these contexts, the use of the adverb, i.e. 'really', signals that the person at issue has propositional knowledge about the subject matter, which is significant in some

contextually salient way, e.g. others might not have this knowledge. In such cases, to know what X really is does not necessarily mean that one knows the essence or some essential properties of X.

In general, 'knowing what X really is' means something like the following: one knows some proposition P about X and P is significant in some contextually salient way, contrasting with propositions one already knows or others know about X. P could pick out a variety of properties of X. For instance, in uttering the sentence 'Elijah knows what a diamond really is', the speaker might have in mind that Elijah has knowledge about (i) a diamond's chemical composition (its essence); (ii) how much it costs (its monetary value); (iii) where it comes from (its origin); (iv) what it could be made to do mechanically (its industrial use); (v) what cultural meaning it carries (its symbolism); and so forth.

Although propositional knowledge P, indicated by the utterance 'someone knows what X really is', can be heterogeneous, the kinds of properties of X that P could be latching onto can be sorted into two salient categories. In many contexts, we are particularly interested in what the underlying essence or nature of something is. This is no surprise since the notion of essence plausibly plays an important role in our everyday thinking. So in hearing the phrase 'S knows what X really is', we might tend to interpret it to mean that S knows some proposition P where P describes the essence or an essential property of X. Call this 'the essential reading' of 'know what X really is'. In other contexts, we are merely interested in some specific property of X, signalled by the relevant context, which has nothing to do with X's essence. So, alternatively, P could be a proposition that

describes a contextually salient but non-essential property of X. Call this a 'non-essential reading' of 'know what X really is'.

Here, it is worth saying something more about the essential reading of 'know what X really is'. Recall (from chapter 4, section 3) that Lewis (1995: 141-2) distinguishes knowing what something is in what he calls 'an uncommonly demanding and literal sense of "knowing what"', which requires the subject to know the essence of the thing in question, from knowing what something is in the everyday, non-demanding sense. Indeed, the demanding sense of 'knowing what' is a familiar interpretation of 'knowing what something is'. In interpreting Socrates on what it means to know what something is, Gail Fine (2014: 35) writes, 'On a familiar view, the claim is that to know what something is, is to know its essence; and to know what something is like (or, equivalently, to know anything further about the thing) is to know its nonessential properties.' My point in this subsection is that a speaker who says that someone knows what something really is might best be interpreted as expressing Lewis' 'uncommonly demanding and literal sense of "knowing what"'.

Context usually makes it clear which kind of reading – the essential reading or a non-essential reading – is in force. Consider the following two utterances:

- (21) Eloise knows how to replicate a diamond because she knows what a diamond really is – a diamond is in fact a solid form of pure carbon.
- (22) Most people in the West don't know what a diamond really is – it is really a symbol of war in Africa.

It is clear that the essential reading is in force in (21), and a non-essential reading is in force in (22). In the absence of contextual information, one might interpret the phrase 'S knows what X really is' on either way. With this clarification regarding the construction 'know what X really is', let's return to sentences (1) to (3).

3.3. EXPLAINING THE DATA

Consider sentence (1), which does not sound odd at all:

- (1) I know what an emerald looks like, but I don't know what an emerald *really* is.

The word 'but' signals a contrast between certain facts about X which the subject knows, and other facts about X which the subject does not know. In (1), I know some propositions about emeralds, namely, 'An emerald looks thus-and-so', but there is some other kind of propositional knowledge about emeralds which I do not have. The lack of such knowledge is indicated by the clause 'I don't know what an emerald *really* is'.

As we just saw, there are two kinds of reading of 'know what X really is': (a) the essential reading, and (b) non-essential readings. Let's consider these two kinds of reading in turn with respect to sentence (1). On the essential reading, in saying that 'I don't know what an emerald really is', I take myself to lack knowledge about the scientific definition or underlying essence of emeralds. With this meaning intended, one can easily imagine sentences like (1) being uttered in a science class. If one has this kind of context in mind in interpreting (1), one would not find (1) odd.

But of course, there are also non-essential readings of ‘know what X really is’. One can easily build a scenario where ‘know what X really is’ is used in a way that is not about essences. For instance, imagine someone putting out ‘fake news’ about emeralds:

- (23) Everyone knows what an emerald looks like – it looks beautiful
and expensive – but they don’t know what an emerald *really* is – it
is poisonous.

Here, the adverb signals that there is some contextually salient property of emeralds that everyone is ignorant of, i.e. being poisonous. It could be the case that one doesn’t find (1) odd because one has this kind of scenario in mind.¹¹ So, it seems that (1) does not sound odd on either the essential or non-essential readings.

Let’s now consider (2) which might sound odd to some but not odd to others:

- (2) I know what an itch feels like, but I don’t know what an itch *really*
is.

In (2), the contrastive stress on ‘feels like’ and ‘really is’ indicates that some knowledge about an itch has already been mentioned or can be assumed (e.g. ‘An itch feels like thus-and-so’), but the speaker wishes to communicate that some other kind of knowledge about itches is lacking.

Let’s consider the two possible kinds of reading of ‘know what X really is’, and this time a non-essential reading first. On non-essential readings, the adverb in (2) is

¹¹ Cases like (23) abound. Consider a more mundane example: ‘I knew what an emerald looked like – it looks like a bright green transparent stone, but I didn’t know what an emerald *really* was – it is in fact a rare and expensive gemstone’.

interpreted as signalling some contextually salient but non-essential property. Imagine Diogenes the Cynic preaches to the Athenians:

- (24) You know what an itch feels like, but you don't know what an itch *really* is. Itches are there to remind us that we are merely animals just like dogs.

Here, Diogenes is claiming that itches have a salient property of being a particular kind of reminder to us, but this property is intuitively not an essential property of itches. If one has scenarios like that of (24) in mind when reading (2), then (2) would not come out as odd at all.

On the essential reading, where the adverb is interpreted as signalling essence talk, I suspect that (2) would sound odd at least to some.¹² But perhaps, it might not sound so odd to others. Compare (2) to (25):

- (25) I know what depression feels like, but I don't know what depression *really* is.

I suspect that (25) would not sound odd to many on the essential reading. It does not sound odd because we now tend to treat depression as a physical state of the brain (e.g. Palazitou 2012; Pandya et al. 2012), and as a result, there is a distinction to be drawn between feeling depressed and the actual physical condition which is depression. Just as one might take depression to be defined as a type of physical state of the brain, one might,

¹² Itches are commonly thought of as sensations or feelings. (2) in this case sounds odd just like (3) does, and its oddness is to be explained in the same way as the oddness in (3), as we shall see.

similarly, take itches to be a type of physical state of the body. If so, then there is a distinction to be drawn between *an itch*, which is a bodily phenomenon, and *the experience of an itch*, which is a mental phenomenon. On this conception, itches themselves are not phenomenal experiences. On the other hand, itchy experiences are certainly phenomenal experiences.¹³

Now, let's consider (3), which sounds, at least to many of us, very odd:

- (3) I know what an itch feels like, but I don't know what the feeling of an itch *really* is.

At the beginning of section 3, I argued that the oddness in (3) is plausibly due to the fact that we are interpreting the second conjunct in a particular way which results in a tension between the two conjuncts. So we have premise (i) as below:

- (i) There is some reading of 'know what X really is' which allows (3) to be rendered odd.

In subsection 3.2, we also established premise (ii):

- (ii) There are two kinds of reading of 'know what X really is': (a) the essential reading; (b) non-essential readings.

I shall now argue for premise (iii):

¹³ In a similar vein, some have argued that pains are bodily states (Reuter 2011; Massin 2017; Reuter, Dustin, and Sytsma 2014; Reuter and Sytsma 2018; Reuter, Sienhold, and Sytsma 2019). If so, then there is a difference between *pains*, which are bodily states, and *pain experiences*, which are phenomenal experiences.

- (iii) A non-essential reading of 'know what X really is' in (3) does not render the sentence odd.

With a bit of imagination, one can construct a scenario which would make a non-essential reading appropriate, in which case (3) would not be rendered odd. Consider our example of Diogenes again. Imagine a particular Athenian, upon hearing Diogenes' sermon on itches, thinks to himself:

- (26) I knew what an itch felt like, but I didn't know what the feeling of an itch *really* was. Diogenes is right. The feeling of an itch is there to remind us that we are merely animals like dogs.

I suspect that most of us would not find (26) odd on this non-essential reading. What this means is that we would not find (3) odd if we were to consider contexts like that in (26). We can then say that there are *some* contexts in which a non-essential reading of 'know what X really is' does not render (3) odd. This of course does not immediately warrant the conclusion that there are *no* contexts in which a non-essential reading is clearly in force and (3) sounds odd. But although the kind of knowledge about non-essential aspects of the feeling of an itch indicated by the second conjunct 'I don't know what the feeling of pain *really* is' might be heterogenous,¹⁴ there is no reason to think that

¹⁴ Consider some other non-essential features of the feeling of an itch which the relevant kind of propositional knowledge might pick out: I knew what an itch felt like, but I didn't know what the feeling of an itch really was. It

- (i) was a symptom of my anxiety;
- (ii) was what destroyed my confidence;
- (iii) was a constant battle;
- (iv) was what led me to become a doctor and specialize in skin diseases;
- (v) was something that was revered by the superstitious villagers;
- (vi) was something that was used by the monks to practice self-control; etc.

(3) would be rendered odd in a particular context in which the non-essential reading is in force.¹⁵ After all, there is nothing contradictory about claiming that one knows what an itch feels like, i.e. ‘An itch feels like thus-and-so’, and then going on to state that one is ignorant about some specific non-essential feature associated with *the feeling of an itch*, whatever that non-essential feature turns out to be. We might then plausibly conclude that a non-essential reading of ‘know what X really is’ simply does not render (3) odd, i.e. premise (iii).

Now, we know that there is some reading of ‘know what X really is’ that allows (3) to be rendered odd (i.e. premise (i)). Since knowing what X really is amounts to knowing either (part of) the essence of X or some non-essential property of X which is contextually significant in some way (i.e. premise (ii)), and since these non-essential readings do not render (3) odd (i.e. premise (iii)), one may then conjecture that the contexts in which (3) is rendered odd are contexts where we are considering the sentence under the essential reading. So we have Premise I:

(3) would not sound odd if one had one of these scenarios in mind.

One can also imagine a conversational context where the present tense, instead of the past tense, is used, i.e. as in (3). For instance, when being asked by Diogenes what the feeling of an itch really is but having failed to provide a satisfactory answer, our Athenian, who defers to Diogenes’ authority, might utter: ‘I guess I know what an itch feels like, but I don’t know what the feeling of an itch really is’, meaning that there is some propositional knowledge about the feeling of an itch which Diogenes knows but the speaker fails to know.

¹⁵ Note that an utterance might sound odd because it is socially or pragmatically inappropriate given the conversational context. Here we are not concerned with the kind of oddness that is due to pragmatic reasons. To say that a sentence is odd here is to say that it sounds contradictory or somewhat inconsistent.

(Premise I) The essential reading of (3) allows the sentence to be rendered odd.

Note that the relevant conclusion here is *not* that sentence (3) with the essential reading of ‘what the feeling of an itch *really* is’ invariably sound odd. In some contexts, (3) might not sound odd when the essential reading is in force. Imagine that our inquisitive Athenian, preoccupied with the feeling of an itch, travels through time to the present and hears a captivating lecture on neuroscience and the mind-body problem from a renowned physicalist. Persuaded by the lecturer, our Athenian might think to himself:

(27) I know what an itch feels like, but I don’t know what the feeling of an itch *really* is. When we figure out the neural basis of consciousness, we shall be able to find out what the feeling of an itch *really* is.

Indeed, a scientifically minded person might not find (27) odd where the essential reading is in force.¹⁶ Such a person might not find (3) odd either. But even if we grant that (3) is not always rendered odd on the essential reading, this does not undermine Premise I. Premise I only says that the essential reading *allows* (3) to be rendered odd. This is compatible with maintaining that there could be contexts, such as (27), in which the essential reading is in force but (3) is not odd. The thought here is that there are everyday contexts in which the essential reading of (3) renders the sentence odd, and this reading might not render the sentence odd if the interpreter was *in a*

¹⁶ Sentence (27) does not sound odd if we are interpreting it on a non-essential reading, in which case we think that neuroscience might be able to tell us some important, non-essential feature of the feeling of an itch, i.e. its neural basis.

scientific/philosophical context. So just because (3) might not be rendered odd in some contexts where the essential reading is in force does not undermine Premise I, nor the target conclusion to be established in the next section, that revelation is part of the *ordinary* conception of experience.

I will address an objection against Premise I in section 5. In the following section, I shall argue for Premise II.

4. PREMISE II

Recall Premise II:

(Premise II) The fact that the essential reading of (3) allows the sentence
to be rendered odd is best explained by appealing to the
hypothesis that the thesis of revelation is part of our
ordinary conception of experience.

If the thesis of revelation is part of our ordinary conception of experience, then it should be no surprise that we implicitly draw on the thesis in thinking and making judgements about experience, including judgements about whether particular sentences about experiences are odd. Let's consider how the hypothesis that the thesis of revelation is part of our ordinary conception of experience provides an explanation for the oddness found in (3) on the essential reading of 'know what X really is'. Consider (3) again:

(3) I know what an itch feels like, but I don't know what the feeling
of an itch *really* is.

In English, the word ‘feeling’ is sometimes understood as designating a phenomenal type, i.e. referring to a specific phenomenal property, as in ‘the feeling of pain’, ‘the feeling of loneliness’, ‘the feeling of being in love’, etc.¹⁷ The phrase ‘the feeling of an itch’, in this case, is thus naturally understood as designating a phenomenal experience-type, i.e. the phenomenal property or phenomenal character of an itchy experience – the itchiness of an itch).¹⁸

Given the first conjunct of (3), ‘I know what an itch feels like’, one would expect the subject to have experienced a token experience of an itch with phenomenal property Q (where Q is the feeling of an itch). Given the thesis of revelation, one would expect the subject to know a proposition ‘Q is X’ where the predicate ‘X’ captures the essence of Q (e.g. ‘The feeling of an itch is thus-and-so’). For the subject to continue with the second conjunct – ‘but I don’t know what the feeling of an itch really is’ – makes the sentence sound contradictory and hence odd. In contrast, (1) is not odd on the essential reading, because it is not part of our ordinary conception of things like emeralds that we know their essences on the basis of ordinary encounters with them. So the hypothesis that the thesis of revelation is part of the ordinary conception of experience explains very well

¹⁷ Similarly, the English word ‘experience’ is often used to simply mean ‘phenomenal character’ or ‘phenomenal property’ (Stoljar 2006: 20; see also chapter 1, footnote 5).

¹⁸ Alternatively, (3) can be formulated explicitly in terms of the phenomenal property *itchiness*:

(3’) I know what an itch feels like, but I don’t know what the itchiness of an itch really is.

I think this sentence is equally odd, even setting aside the fact that ‘itchiness of an itch’ is a somewhat unnatural expression.

why the essential reading of (3) allows the sentence to be rendered odd. In the absence of a better candidate explanation, we shall provisionally take it to be the best explanation.

Let's sum up our argument so far. In the last section, we established that:

(Premise I) There is some reading of (3) which allows the sentence to be rendered odd, i.e. the essential reading.

In the absence of a better candidate explanation for the fact reported in Premise I, we can (at least provisionally) conclude Premise II:

(Premise II) The fact that the essential reading of (3) allows the sentence to be rendered odd is best explained by appealing to the hypothesis that the thesis of revelation is part of our ordinary conception of experience.

By inference to the best explanation, we can conclude:

(Conclusion) The thesis of revelation is part of our ordinary conception of experience.

In the next section, I address objections against this argument.

5. OBJECTIONS

Potential objections against the above argument can be levelled against either Premise I or Premise II. I will first address a potential objection against Premise II and then consider a potential objection against Premise I.

Objection 1: There is a better explanation for the oddness of sentences like (3) on the essential reading. The thesis of revelation belongs to a particular conception of experience invented by philosophers, but which is not particularly hard to latch onto once one starts thinking philosophically about experience. Once it is latched onto, one would find sentences like (3) odd on the essential reading. But this is not to say that the best explanation for the latter is the hypothesis that revelation is part of the *ordinary* conception of experience. On this alternative explanation, the oddness of (3) on the essential reading surfaces in the philosophical context, but not in everyday contexts.

Reply: This putative alternative explanation leaves unanswered just why it is easy for the folk to latch onto and draw on this conception of experience invented by philosophers, which incorporates the thesis of revelation. Indeed, there are many conceptions of ordinary phenomena invented by philosophers that are not particularly easy for the folk to arrive at on their own, even if they start to think about these phenomena philosophically, e.g. the nihilist conception of ordinary objects according to which the latter do not exist; the panpsychist conception of experience, on which experience is ubiquitous; the idealist conception of reality, on which all things are fundamentally mental, and so on. So, the claim that the conception of experience that incorporates the thesis of revelation is invented by philosophers leaves it unclear why it is a conception that the folk can easily latch onto when they start thinking about experience philosophically. In contrast, if revelation is part of the *ordinary* conception of experience, then there is no further question as to why, and indeed no surprise that the folk would draw on this conception of experience and the thesis of revelation in judging

whether sentences like (3) are odd.¹⁹ The hypothesis that revelation is part of the ordinary conception of experience offers a much more straightforward explanation for the oddness in (3). Objection 1 thus fails to problematise Premise II.

Objection 2: It is not the case that the essential reading allows sentence (3) to be rendered odd. (3) sounds odd because we have confused it with the following sentence where the adverb is positioned before ‘know’ rather than before ‘is’:

(28) I know what an itch feels like, but I don’t *really* know what the feeling of an itch is.

There is a reading of (28) which renders the sentence very odd. According to this reading, ‘really’ means ‘actually’ or ‘in fact’; ‘S doesn’t *really* know what X is’ entails ‘S doesn’t know what X is’. On this reading, (28) sounds contradictory, because it entails something like ‘I know what the feeling of an itch is and I don’t know what the feeling of an itch is’, which is contradictory at the level of surface grammar.

Granted the confusion of ‘know what X *really* is’ with ‘*really* know what X is’, it seems that this proposal also explains nicely why we don’t find the rest of the sentences as odd as (3). According to this proposal, if we are confusing (3) with (28), then we must also be confusing (1) with (29), and (2) with (30):

¹⁹ This is not to say that we always draw on our tacit understanding of revelation in judging sentences like (3). In some contexts, the tendency to appeal to one’s tacit understanding of revelation can be trumped by one’s commitment to, say, a reductionist physicalist worldview. But this is compatible with revelation being part of the *ordinary* conception of experience. More on this issue in chapter 7.

(29) I know what an emerald looks like, but I don't *really* know
what an emerald is.

(30) I know what an itch feels like, but I don't *really* know what an
itch is.

(29) entails something like 'I know what the look of an emerald is, but I don't know what an emerald is', which seems acceptable. (30) entails something like 'I know what the feeling of an itch is, but I don't know what an itch is', which is certainly not a contradiction at the level of surface grammar.

So, the objection goes: in considering sentences (1) to (3), we are systematically confused with respect to the position of 'really' in the second conjunct of each of the sentences, i.e. confusing 'know what X *really* is' with '*really* know what X is'. On this confused reading, (3) would come out as contradictory, whereas the rest of the sentences either don't sound odd at all (i.e. (1)), or don't sound odd to the same extent (i.e. (2)). On this alternative explanation for the oddity in (3), the sentence is odd not because we are interpreting it according to the essential reading; it is odd because we are confusing it with (28). So, Premise I is false.

Reply: It is problematic to assume that competent English speakers cannot tell the difference between 'S doesn't *really* know what X is' and 'S doesn't know what X *really* is', and the relevant semantic differences. Consider this scenario. Imagine that your friends are talking about tourmaline. Having never heard of tourmaline and wanting to confess your ignorance, you would naturally say: 'I don't *really* know what tourmaline is' or 'I don't know what tourmaline is'. It would seem odd for you to say 'I don't know

what tourmaline *really* is', if you in fact have never heard of tourmaline and do not know anything about tourmaline. Consider another example. Imagine that you want to try the best restaurant in London. Knowing that your friend Jamie is a foodie, you ask Jamie what he thinks the best restaurant is. Jamie replies: 'I don't *really* know what the best restaurant is'. This might come as a surprise to you insofar as it contradicts your expectation that Jamie would have an opinion about what the best restaurant in London is. Imagine that Jamie instead replies: 'I don't know what the best restaurant *really* is'. This latter reply hardly makes sense in the context. If Jamie replies to you this way, you might start wondering whether he has even understood your simple question in the first place.

Given the aforementioned examples, ordinary English speakers seem to be perfectly competent at deciphering the differences between cases where 'really' modifies the verb 'know', and cases where 'really' is used to signal a particular contextually-salient aspect of a thing at issue. Objection 2 thus fails to problematise Premise I.

6. CHAPTER CONCLUSION

In this chapter, I argued that the thesis of revelation is intuitive because it is part of our ordinary conception of experience. In putting forward my argument, I considered ordinary discourse on essence and experience. After eliciting certain linguistic intuitions that we have about experience, I contended that the best explanation for such intuitions is the hypothesis that revelation is part of the ordinary conception of experience. By inference to the best explanation, one can conclude that the conception of experience that

incorporates revelation is indeed the ordinary one. Following from this, we also have a reason to think that the thesis of revelation is an intuitive claim. In the next chapter I put forward an explanation for the intuition of dualism in terms of the intuitiveness of revelation in the sense at issue.

Chapter 7

EXPLAINING THE INTUITION OF DUALISM: REVELATION

1. INTRODUCTION

Let's take stock of what we have achieved so far in Part II. In chapter 4, we clarified the thesis of revelation. In chapter 5, we spelt out the incompatibility between revelation and standard versions of physicalism, which claim that phenomenal properties have physical or functional essences. In chapter 6, we argued that revelation is intuitive in the sense that it is part of the ordinary conception of experience. In this chapter, we will return to the question left unanswered at the end of Part I, the question of what explains the intuition of dualism, the belief or disposition to believe that consciousness is nonphysical. In chapter 3, we reviewed a number of proposals for explaining the intuition of dualism, but they all failed to be satisfactory for one reason or another. In this chapter, we consider a proposal for explaining the intuition of dualism which appeals to the intuitiveness of the thesis of revelation (section 2). I shall also defend this proposal from potential objections (section 3).

If there is a plausible explanation of the intuition of dualism based on the intuitiveness of revelation, one might wonder whether, as with respect to the thesis of revelation regarding phenomenal properties, we also intuitively hold a similar thesis of revelation in a different domain. If so, then we might anticipate a similar intuition of distinctness, like the intuition of dualism, holding some alleged properties to be intuitively distinct from physical properties. One particular domain that stands out is

colour. As already mentioned, there is also a thesis of revelation regarding colour properties (Johnston 1992). In section 4 of this chapter, I compare the thesis of revelation about colour properties with the thesis of revelation regarding phenomenal properties. In addition, I consider the question of whether there is an intuition that colour is nonphysical.

If there is a plausible explanation of the intuition of dualism based on the intuitiveness of revelation, one might also wonder why revelation is intuitive in the first place, that is, why it gets to be part of our folk conception of experience. In section 5, I spell out an explanation for the intuitiveness of revelation that depends on the truth of the thesis of revelation itself.

The issue of what explains the intuitiveness of revelation may be of particular interest to defenders of physicalism. If physicalists can provide a satisfactory explanation that is compatible with physicalism, then they can maintain that revelation is intuitive in the sense of being part of our ordinary conception of experience while not accepting revelation as true. In section 6, I look at a notable physicalist proposal for explaining the intuitiveness of revelation, which appeals to the hypothesis that there might be a systematic introspective misrepresentation of the natures of phenomenal properties. I contend that this particular proposal is unsuccessful. Section 7 concludes the chapter and Part II of the dissertation.

2. THE EXPLANATION

Building on discussions in preceding chapters, the intuition of dualism, that is, our belief

or disposition to believe that phenomenal properties are not physical properties,¹ can be explained by the intuitiveness of the thesis of revelation in the following way. Our intuition of dualism has an inferential aetiology that can be traced to our tacit understanding of the thesis of revelation, which is part of our ordinary conception of experience, and a tacit appreciation of the entailment from the thesis of revelation to the claim that phenomenal properties are not physical properties. In this sense, we can also say that the thesis of revelation is the rationale that underpins our intuition of dualism.

Given what has just been said, one might immediately worry that intuitions, including the intuition of dualism, are not supposed to be inferential. There are two senses in which an intuition is said to be non-inferential – a psychological sense and an epistemic sense (Jackson 2018). Understood in the psychological sense, intuitions are non-inferential because they are not the results of inferences from other doxastic states (Jackson 2018). More commonly held is the view that intuitions are non-inferential in the epistemological sense, whereby they are either non-inferentially justified or provide immediate justification for certain beliefs (Bealer 1996, fn20; Goldman and Pust 1998; Chalmers 2014). It is the psychological sense of non-inferentiality that might seem to be in tension with my claim that the intuition of revelation has an inferential aetiology. But it is worth pointing out that in stating that intuitions are non-inferential in the psychological sense, philosophers often mean that they are not formed as the result of the subject *consciously* inferring them from premises with contents of other doxastic states. For instance, Joel Pust (2000: 30) makes this explicit: ‘many philosophers seem to

¹ Here, I am using the term ‘physical’ to include ‘functional’.

use “intuition” to refer to a kind of spontaneous or non-inferential judgement or belief ... Intuitions are not the result of conscious inference’ (see also Weinberg 2007: 318). It seems very plausible that an intuition such as the intuition of dualism could still have an inferential or inference-like aetiology which the subject is unconscious of or has no introspective access to.

Indeed, many of our contentful mental states have inference-like aetiologies which we are not consciously aware of, but which we appeal to in explaining a subject’s judgements or behaviours in general. Consider someone who can fluently read Russian script aloud. In doing so, the speaker draws on her knowledge of letter-sound rules, i.e. rules that go from certain orthographic representations to corresponding phonological representations (see also Davies 2015). She does not consciously infer from these rules in reading. If she needed to do so, she would not be able to read fluently. Implicit knowledge of these rules explains the speaker’s correct reading of a particular Cyrillic letter string as well as her general ability to read Russian scripts.

Consider also Peacocke’s (2008: 114) example of how someone who has not yet learned the relevant truth tables can come to appreciate the validity of the primitive axiom schema: $A \rightarrow (A \text{ or } B)$; or the primitive inference rule: ‘From A, a conclusion of the form “A or B” can be inferred’. Peacocke (2008: 115) suggests that the thinker’s acceptance of the axiom or rule as valid involves a simulation exercise. For instance, she first imagines the case that A is true and B is false and, drawing on her understanding of alternation, comes to evaluate ‘A or B’ as true. She then imagines the case that A is true and B true, and so on through all the cases. Eventually, she is ‘in a position to accept

rationally that there will be no cases in which the antecedent, or premise, [A] is true, and the consequent, or conclusion, ['A or B'] is false [and thus to] accept rationally the axiom or rule as valid' (Peacocke 2008: 116).

In the starting case, Peacocke explains our thinker's judgement that 'A or B' is true in terms of the simulated truth-values in the case (A is true and B is false), plus her *understanding* of the connective 'or'. Crucially, for Peacocke (2008: 116-7), the thinker's understanding of the connective 'or' involves her possession of an implicit conception, i.e. a state of tacit knowledge, 'with the following content: that any sentence of the form "A or B" is true if and only if either A is true or B is true'. Peacocke gives a similar account of the implicit conception that is involved in the thinker's grasp of the concept of alternation. The thinker's pattern of judgements about various cases, as well as her eventual appreciation that the logical axiom or rule is valid, is understanding-based. It draws on her understanding of the connective 'or' or her grasp of the concept of alternation, and thus on her possession of an implicit conception. But this does not mean that she needs to make a conscious inference from a premise with the content of this implicit conception. As Peacocke (2008: 117) puts it, the implicit conception is 'influential in the thinker's evaluation of alternations' but the thinker 'need not have any explicit knowledge of its content'.

I think our intuition of dualism works in a similar way. Our judgement that phenomenal properties are not physical properties draws on our tacit or implicit understanding of the thesis of revelation (which is part of our ordinary conception of experience) and on a tacit or implicit appreciation of the logical relation between the

thesis of revelation and the claim that phenomenal properties are not physical properties. This does not mean that we need to reason consciously and meticulously from the premise of the thesis of revelation to the conclusion that phenomenal properties are not physical properties, nor do we need to explicitly entertain the thesis of revelation itself.

Of course, this does not exclude the possibility that upon suitable prompting, we can make explicit our implicit understanding of the thesis of revelation, just as upon suitable enquiry, it is possible that our Russian speaker can explicitly state the letter-sound rules she appeals to, and that the thinker in Peacocke's case can make explicit her implicit conception of the connective 'or'. So in an important respect, our tacit understanding of the thesis of revelation is different from Chomskyan tacit knowledge of linguistic rules, which explains our judgements of whether certain sentences are grammatical. One might think that in the latter case, since the grammatical notions involved in linguistic rules are technical notions, explicit statements of the linguistic rules cannot be elicited from the subject even upon suitable prompting.

One might further add that our intuition of dualism typically arises when we reflect on particular instances of phenomenal experiences that we are having or calling to mind. So, the intuition of dualism can be conceived as the output of an inferential or inference-like transition from having or recalling such experiences. We can further conceive this transition as involving two sub-transitions. In the first place, there is a transition from having or recalling a mental state with a *particular* phenomenal property Q_0 , to a representational state with the following content:

[R1] Q_0 is not a physical property.

This transition is facilitated by our understanding of the thesis of revelation and an appreciation of the entailment from the thesis of revelation to the claim that Q_0 is not a physical property. This entailment can be formulated in a similar way to our formulation of the revelation argument against physicalism in section 5 of chapter 5:

- (1) If S has an experience-token with quale Q_0 , then S is in a position to know *de dicto* that ' Q_0 is X_0 ', where the predicate ' X_0 ' captures the essence of Q_0 .
- (2) If Q_0 is a physical (including functional) property, then Q_0 has a physical essence.
- (3) If Q_0 has a physical essence,² then by having an experience-token with quale Q_0 , S should be in a position to know *de dicto* that ' Q_0 is X_0 ', where ' X_0 ' is a physical predicate which captures the essence of Q_0 .
- (4) It is not true that by having an experience-token with quale Q_0 , S is in a position to know *de dicto* that ' Q_0 is X_0 ', where ' X_0 ' is a physical predicate which captures the essence of Q_0 .
- (5) Q_0 is not a physical property.

The above is an argument from the thesis of revelation formulated with respect to a particular phenomenal property, i.e. Q_0 , to the claim that Q_0 is not a physical property, i.e. [R1]. I have argued that revelation is intuitive in the sense of being part of our

² Here 'physical' is again used in the sense that includes 'the functional'.

ordinary conception of experience. So it is most likely that information regarding revelation can be drawn on – accessed and used – when we think about a particular phenomenal property of our experience.

In the second place, there is a transition from [R1] to a representational state with a more generalised content:

[R2] Phenomenal properties are not physical properties.

The thesis of revelation does not make a specific claim about a specific phenomenal property. It is a general rule that can be deployed when we think about any phenomenal property. So, in reaching the conclusion that Q_0 is not a physical property, one can appreciate that the conclusion is not drawn from specific information about Q_0 . The conclusion [R1] depends only on Q_0 falling within the range of the variable 'Q', which is restricted to phenomenal properties. So, by universal generalisation, what goes for Q_0 goes for any phenomenal property, and we have the conclusion [R2]. Thus, this second transition from [R1] to [R2] involves our tacit appreciation of the rule of universal generalisation.

The proposed explanation of the intuition of dualism in terms of the intuitiveness of the thesis of revelation can be further elucidated by comparing phenomenal types with other types such as *natural kinds*. Let's agree that the property of being water is essentially the property of being H_2O . It is not part of our ordinary conception of water that we are in a position to know the essence of water in having an ordinary experience of water. As a result, when scientists inform us that water is H_2O , we do not resist the

identification or find it puzzling.³ Were we to think that we are in a position to know the essence of water in having an experience of water, we would have a persistent intuition of distinctness – like our intuition of dualism – that water is not H₂O, or that the property of being water is not a chemical property.

It is also interesting to compare phenomenal types with *functional kinds*. Consider the functional property of being furniture. I think it is part of our ordinary conception of furniture that in knowing the function of furniture, we know what it is to be furniture, i.e. being furniture is essentially being movable articles in a room which make the room suitable for living or working in. Imagine that some expert comes along and tells us that furniture is not what we think it is, and all pieces of furniture are, say, religious artefacts. I think we would not take the expert to be making a statement about what furniture *essentially* is, but merely stating a contingent but interesting fact about furniture that most people don't know. But if the expert insists that she is making a claim about what the essence of being furniture is, e.g. being furniture is being artefacts that glorify God in a particular way, then we would naturally give her an incredulous stare. This is similar to the case of consciousness, where we have the intuition that phenomenal properties are not what physicalists claim they are.

So far, we have seen that the intuition of dualism may plausibly be explained by the intuitiveness of the thesis of revelation. In the next section, I address three objections against the proposed explanation.

³ One might feel an initial sense of puzzlement when confronted with the identity claim 'Water is H₂O', since one cannot infer from an ordinary perceptual experience of water that water is H₂O. But this puzzlement does not seem to be persistent.

3. OBJECTIONS

Objection 1: In his brief discussion of the proposal that the intuition of dualism might be explained in terms of the intuitiveness of the thesis of revelation, Papineau (2011: 17) raises the concern that such an explanation ‘assumes an overly sophisticated level of reflection about the mind-body issue on the part of ordinary thinkers’.⁴

Reply: I think the intuition of dualism may indeed require some level of reflection, just as one might need to reflect before rationally accepting the logical axiom $A \rightarrow (A \text{ or } B)$ (or the corresponding rule of inference). In order to have the intuition that consciousness is nonphysical, one would need to first grasp the concept of consciousness, which in turn requires one to think about our experiences in a particular way, that is, in terms of their what-it-is-likeness or phenomenal characters. But as I have clarified in the last section, our dualistic intuition does not assume an overly sophisticated level of reflection if the relevant sophistication is understood in the sense that the subject needs to reason explicitly from the thesis of revelation to the conclusion that phenomenal properties are not physical properties. As we saw in the section 2, the relevant entailment from the thesis of revelation to the claim that a particular quale is not physical and also to the general claim that phenomenal properties are not physical properties requires careful spelling-out if we are to make it clear for a philosophical audience. But this is not to say that the entailment cannot be tacitly or implicitly appreciated by someone who lacks philosophical training and lacks an explicit understanding of the thesis of revelation. For

⁴ In a forthcoming paper, Papineau acknowledges that revelation is ‘a highly intuitive idea’, and that many people ‘embrace dualism for that reason’.

instance, in considering the phenomenal property of phenomenal blue upon seeing a pure blue sky, one might easily come to appreciate, drawing on one's implicit understanding of the thesis of revelation, that the phenomenal property is not some physical property cashed out in neurophysiological terms, or some functional or causal-role property spelt out in commonsense or technical vocabulary.

Objection 2: In an unpublished paper, Andrew Melnyk argues that what he calls 'the sense of incredibility' – that it is incredible to take phenomenal properties to be physical properties – explains the 'feeling' of revelation. He writes:

The sense of incredibility may well be what underlies the feeling that introspection reveals the complete essence of phenomenal properties. For if introspection doesn't reveal the complete essence of a given phenomenal property, then there must be some true informative identity claim of the form 'That (introspected) property = F', where 'F' is a term from, say, neurophysiology. But it's precisely claims of this form that one finds utterly incredible when one experiences the sense of incredibility.

Someone like Melnyk might say that it is not that the intuitiveness of the thesis of revelation explains the intuition of dualism, but, rather, the other way around: the intuition of dualism explains why we hold or are inclined to hold the thesis of revelation. So, the proposed explanation fails to be satisfactory because it gets the explanatory order wrong.

Reply: Even if the intuition of dualism constitutes an explanation for a 'feeling' or intuition of revelation in some cases, this does not undermine my proposal, because the intuition of dualism and the intuition of revelation might be mutually supporting. Consider the following scenario. A subject has an explicit belief that phenomenal

properties are not physical properties. According to my proposed explanation, this belief can very well be due to the intuitiveness of revelation, the subject's drawing on her implicit understanding of revelation (which is part of our ordinary conception of experience), even if she has not explicitly thought about the thesis. Now the belief that phenomenal properties are not physical properties might lead the subject to explicitly appreciate the thesis of revelation, just as Melnyk suggests.⁵ But none of this problematises my proposal that the intuition of dualism itself can be explained by the intuitiveness of revelation.

Objection 3: The proposed explanation does not explain the *persistence* of the intuition of dualism in scientific and philosophical contexts. Recall that in the last chapter, we considered the following sentence:

(27) I know what an itch feels like, but I don't know what the feeling of an itch *really* is. When we figure out the neural basis of consciousness, we shall be able to find out what the feeling of an itch *really* is.

The point being made there was that when we raise scientific and theoretical considerations regarding the mind and the brain, we might become less averse to the idea that *we don't know* what the feeling of an itch really is, just by experiencing an itch.

⁵ Note that there is an entailment from the thesis of revelation to the claim that phenomenal properties are not physical properties, but the reverse is not true. One can maintain a metaphysical position according to which phenomenal properties do not have physical essences but have a kind of nonphysical essence which is hidden from experience. It might very well be the case that the subject in the main text fails to consider this metaphysical view, subsequently forming a belief in the thesis of revelation or finding revelation intuitive.

Once one realises – for example, in a scientific context – that the thesis of revelation could be false, surely one should also cease to be in the grip of the intuition of dualism. But the dualistic intuition is supposed to persist even in scientific contexts.

Reply: Even if, by considering statements like (27), one comes to think or consider that the thesis of revelation could be false, this does not mean that one is no longer influenced by its intuitiveness. That is, the *persistent* intuitiveness of the thesis of revelation can still be the explanation of the *persistent* intuition of dualism. Let's consider the case of knowledge. Sceptics who deny, on philosophical-theoretical grounds, that we have everyday knowledge nonetheless participate in ordinary discourse which presupposes that we have knowledge, and make everyday knowledge attributions. It is reasonable to think that those sceptics are, in spite of their explicit theoretical convictions, still influenced by their implicit pre-theoretical intuition that we have knowledge. In a similar way, those who explicitly deny the thesis of revelation on theoretical grounds might still feel the intuitive force of the thesis, and hence also find dualism intuitive at some level. In this sense, one might say, echoing Papineau (2011), that one can be theoretically convinced that dualism is false and that the thesis of revelation is false, but nevertheless find these claims intuitive.

Let's sum up. As we saw in the last section, the thesis of revelation being part of our implicit conception of experience (i.e. being part of our ordinary conception of experience) gives rise to the intuition of dualism. A subject who possesses this implicit conception might, however, not explicitly accept the thesis of revelation. In discussing Objection 2 in this section, we saw that the intuition of dualism, which can be traced to

our tacit understanding of revelation, might lead to an explicit appreciation of the thesis of revelation. As we also saw in discussing Objection 3, instead of taking a positive explicit view of revelation, one might adopt a negative explicit view of revelation because of the influence of science. But none of these cases problematise the claim that there exists an implicit conception that incorporates the thesis of revelation. The overall situation seems to be that our having a tacit understanding that includes the thesis of revelation is consistent with a subject having no explicit view, or a positive explicit view, or even a negative explicit view of revelation.

Given the proposed explanation for the intuition of dualism, in the next section I consider another domain in which a similar intuition of distinctness might potentially arise.

4. REVELATION ABOUT COLOUR

As already mentioned in chapter 4, in analytic philosophy the term ‘revelation’ also refers to the idea that the essences of colour properties (understood as properties of mind-independent objects) are revealed in colour experiences (Johnston 1992). The thesis of revelation about colour (revelation-c for short) and the thesis of revelation about qualia (revelation-q for short in this section) have a number of parallel features:

- (1) Both can be formulated in similar ways.
- (2) Both are incompatible with standard versions of physicalism.
- (3) Both have been thought to be part of our conception of the relevant domain, about colours and qualia respectively.

These parallel features may suggest that an intuition of distinctness, similar to the intuition of dualism, can potentially arise with respect to colour properties. Whether or not there is a shared, persistent intuition that colour is nonphysical is an issue I will consider shortly. First, let us clarify each of these parallel features in turn.

Revelation-c was notably featured in Mark Johnston's 1992 paper 'How to Speak of the Colors'. Revelation-c is associated with a view about colour, sometimes known as the *simple view* of colour or *primitivism* about colour, according to which colours are simple, irreducible, qualitative properties of physical objects (see Campbell 1993). Using the example of canary yellow, Johnston (1992: 223) defines revelation-c as follows:

The intrinsic nature of canary yellow is fully revealed by a
standard visual experience as of a canary yellow thing.

There have been various proposals to sharpen Johnston's original formulation of revelation-c (Campbell 2005; Byrne and Hilbert 2007; Gert 2008; Allen 2011). Notably, Alex Byrne and David Hilbert (2007) formulate Johnston's idea of revelation-c such that knowing the nature of a colour is knowing a collection of chromatic truths about that colour, e.g. 'Red is more similar to orange than blue', 'Red is not a shade of blue', etc., where these chromatic truths capture the structural features of colour space – how colours are related to one another. However, Byrne and Hilbert's formulation struggles to capture the intuitive sense of revelation-c found in Johnston.⁶ In criticising Byrne and

⁶ Given the distinction between essential properties and necessary properties (see Fine 1994), there is a question of whether structural or relational properties are *essential* properties of colours, e.g. whether the property of *being more similar to orange than blue* is an essential property of colour *yellow*, rather than merely a necessary property. Even if they are essential properties, there is still the issue of whether structural properties alone exhaust the nature of colour space.

Hilbert's formulation of revelation-c, John Campbell (2005: 110) argues that revelation-c is meant to capture the idea that 'colour experience reveals the colour itself to one', and that 'the role of experience is not immediately to provide one with...knowledge of the structure of the colours. It is, rather, to provide one with knowledge...of...the qualitative characters of the colors'. For Campbell (2005: 10), we have knowledge by acquaintance of colour, according to which 'experience of color provides knowledge of the categorical color property intervention on which changes the experiences of observers'.⁷ But knowledge by acquaintance of colour, *à la* Campbell, does not capture the original idea of revelation-c from Johnston which says that colour experiences reveal the *natures* of colour properties.⁸ In general, having acquaintance knowledge of something does not guarantee knowledge of the nature or essence of that thing, e.g. one might have knowledge by acquaintance of water, or of Wendy, without knowing the essence of water, or of Wendy. So cashing out revelation-c as a kind of acquaintance knowledge fails to capture Johnston's original idea.

Here I adopt the idea that structural properties are not essential properties of colours, e.g. it is not part of the (immediate constitutive) *essence* of yellow that it is more similar to orange than blue, although the latter may be a necessary feature of yellow. The same goes for structural properties of qualia, e.g. *being more similar to phenomenal orange than phenomenal blue* is not part of the (immediate constitutive) essence of *phenomenal red*.

⁷ Campbell (2005) calls this 'Transparency' to distinguish it from what Byrne and Hilbert have called 'Revelation'.

⁸ In his 1993 paper 'A Simple View of Colour', Campbell seems to subscribe to Johnston's idea of revelation-c, where he takes the real natures of colour properties to be 'transparent to us' (1993: 258). In his 2005 paper 'Transparency versus Revelation in Color Perception', however, it is unclear that Campbell is committed to the idea that the essence of colour is revealed to us in colour experience. So in formulating the idea that we have knowledge by acquaintance of colours, he may not have in mind the goal of developing Johnston's original idea.

It seems to me that Johnston's original idea of revelation-c can in fact be captured adequately by formulating revelation-c in a similar fashion as we have formulated revelation-q:

(Revelation-c)

By having an experience of a colour property C (where C is a property of mind-independent objects),⁹ one is in a position to know a truth, namely, 'C is Y', where the predicate 'Y' captures the essence of C, although it may be hard to put into words.

This formulation of revelation-c avoids criticisms levelled against Byrne and Hilbert's formulation – knowing the essence of a colour is not knowing how the colour stands in relation to other colours in terms of chromatic similarities and differences. Contrary to Campbell's formulation, furthermore, it captures the original idea from Johnston that revelation-c is about knowing the essence of a given colour rather than just having acquaintance knowledge of the colour.

Let's move on to the second parallel feature: both revelation-q and revelation-c are incompatible with standard versions of physicalism. Standard versions of physicalism take colour properties, understood as properties of mind-independent objects, to have physical or functional essences. There are two realist theories of colour

⁹ Without the qualification in the brackets, revelation-c could collapse into revelation-q supposing one maintains subjectivism about colour, according to which colours are phenomenal properties of our subjective experience. Here I am adopting the common-sense view that colours are properties of mind-independent objects. In this way, revelation-c can be clearly distinguished from revelation-q.

that take colour properties to be such properties: physicalism about colour and dispositionalism. According to ‘physicalism about colour’ (see Jackson 1998; Byrne and Hilbert 2007), colours are mind-independent, complex physical properties of objects, e.g. surface reflectance properties. According to dispositionalism, colours are dispositions in objects to cause certain kinds of experience in the perceiver (see Locke 1689; McGinn 1983; Johnston 1992; Peacocke 1997). Dispositions may be identified with functional role properties (Logue 2016: 221; Cohen 2009: 220). Dispositionalism about colour is thus analogous to role functionalism regarding qualia, and physicalism about colour, realiser functionalism or the identity theory (Sundström 2007; Logue 2016: 221). So according to standard versions of physicalism, colour properties have physical or functional essences – physical/functional descriptions exhaust the natures of colour properties.

The incompatibility between revelation-c and physicalism can be set out in an argument parallel to the argument from revelation-q against physicalism which we saw in chapter 5:

- (i) If S has an experience-token of colour C (where C is a property of mind-independent objects), then S is in a position to know *de dicto* that ‘C is Y’, where the predicate ‘Y’ captures the essence of C.
- (ii) If physicalism is true, then colours have physical/functional essences.
- (iii) If colours have physical/functional essences, then by having an experience-token of colour C, S should be in a position to know

de dicto that 'C is Y', where 'Y' is a physical/functional predicate which captures the essence of C.

- (iv) It is not true that by having an experience-token of colour C, S is in a position to know *de dicto* that 'C is Y', where 'Y' is a physical/functional predicate which captures the essence of C.
- (v) Physicalism is false.

Premise (i) is entailed by the thesis of revelation-c. (ii) states what the essences of colour properties would be if physicalism were true. (iii) follows from (i) and (ii). (iv) should also be unproblematic. In having a colour experience, one is not in a position to know that 'C is Y', where 'Y' is a physical predicate such as 'having surface reflectance property R'. It also seems true that 'Y' cannot be a functional predicate, for the similar reason that 'X', which captures the essence of quale Q, cannot be a functional predicate, as discussed in section 3 of chapter 5.¹⁰ (v) then concludes from (iii) and (iv) that physicalism (that is standard versions of physicalism which take colours to have physical/functional essences) is false.

Let's move on to the third parallel feature: both revelation-q and revelation-c are

¹⁰ A dispositionalist would define colour C by making reference to the colour experience C': C is the disposition to cause experience C' in a perceiver (see Logue 2016: 221). There is a question as to what experience C' is. A commonsense role functionalist would define experience C' with respect to C, i.e. C' is an experience caused by perceiving colour C in standard conditions (see Lewis 1997). But as we already saw in chapter 5, the pair <C, C'> does not individuate colour C from other colours. In order to define colours in a way that individuates them, a dispositionalist would have to define colours in functional terms by appealing to standard experiences of seeing worldly objects, e.g. 'Colour red is the disposition to cause the experience that one would have when one sees a British pillar box'. But one is certainly not in a position to know the latter truth about the colour red nor anything about the British pillar box merely by having an experience of the colour red. So dispositionalism thus understood is incompatible with revelation-c.

thought to be part of our ordinary conception of the relevant domain, about colours and qualia respectively. I have already argued in chapter 6 that revelation-q is part of our ordinary conception of experience. Regarding revelation-c, Johnston takes it to be one of our core beliefs about colour (1992: 223), and part of our ‘salient, “intuitively based” conception of colour’ (1992: 228; see also Gow 2014). Campbell (1993: 258) takes revelation-c (in the original sense from Johnston) to be intuitive and commonsensical.¹¹ So some philosophers have thought that revelation-c is part of our ordinary conception of colour.

Our proposed explanation of the intuition of dualism (in section 2, above) appeals to the intuitiveness of revelation-q. Given the parallel features shared by the two theses of revelation, one might suspect that a similar intuition of distinctness could also arise in the case of colours. Whether or not such an intuition would arise, of course, depends crucially on whether revelation-c is a deep part of our ordinary conception of colour, as philosophers like Campbell and Johnston suggest.

Here, it is worth noting that the idea that colour presents a problem for physicalism is not unfamiliar in the literature and has been voiced by a number of philosophers. One might take this to go some way towards suggesting that there could be an intuition of distinctness regarding colour properties. For instance, Sydney Shoemaker (2003) distinguishes between the subjective explanatory gap and the

¹¹ While it is unclear that Campbell holds revelation-c in his 2005 paper, it seems that he would have endorsed revelation-c in his 1993 paper, as well as the intuitiveness of revelation-c. Although he does not explicitly talk about the (immediate constitutive) essences of colours, he (1993: 258) says that ‘the character’ and ‘the real nature’ of a colour property is transparent to us.

objective explanatory gap. The former, which we discussed in chapter 2, has attracted much attention in contemporary philosophy and is the problem of explaining phenomenal properties of experience in terms of physical properties of the brain. The latter goes back to Galileo, Descartes and Locke (Shoemaker 2003: 254), and is the problem of explaining sensible qualities like colours as we see them in terms of physical properties of mind-independent objects. Alex Byrne (2006), in his paper 'Colour and the Mind-Body Problem', argues persuasively that considerations against physicalism, e.g. Kripke's (1980) conceivability argument, Jackson's (1982, 1986) knowledge argument, Levine's (1983) explanatory gap, can be framed in terms of colours instead of qualia.

Colour primitivists like Campbell (1993, 2005), who combines colour primitivism with naïve realism about perception,¹² contend that the real explanatory gap is in fact located with colours out in the world, not with phenomenal properties in the mind (see also Allen 2017).¹³ In discussing the explanatory gap regarding qualia identified by Levine (1983), Campbell (2005: 112) writes: 'it is, indeed, natural to wonder whether Levine's article did not mislocate the fundamental point of contact between the

¹² According to naïve realism, veridical perceptual experiences are at least partially constituted by mind-independent objects and their properties. Naïve realists explain the phenomenal character of perceptual experience in terms of properties of mind-independent objects.

¹³ Colours, together with sizes, shapes, etc., are often thought of as qualitative properties of mind-independent objects. One might also think that mind-independent objects have qualitative properties such as sounds, smells, textures, temperatures, etc., which we directly perceive according to naïve realism. One could also be a primitivist with respect these properties. In the latter case, sounds, smells, etc. also create explanatory gaps.

Note also that one need not commit to naïve realism about perception in order to hold the view that colours, rather than qualia, provide the real challenge for physicalism. The latter position is also compatible with versions of representationalism about perception (see Shoemaker 2003: 255).

qualitative and the physical. The more fundamental ... problem is to understand the relation between the qualitative color property of the object and its physical basis'. For Campbell (1993: 268), 'the qualitative character of a colour-experience is inherited from the qualitative character of the colour'. Accordingly, phenomenal red, which is a phenomenal property of a colour experience, is, at least partially, constituted and explained by the *sui generis* red property of the perceived object. In Campbell's picture, the bump in the rug for physicalism, i.e. the problem of the explanatory gap, is moved from the mind to the external world.

Of course, the received view in the literature is still that qualia, rather than colours, present the real and sole problem for physicalism. In a recent paper, Chalmers (2018: 27) claims that there is 'a crucial disanalogy between the representation of colours and phenomenal properties'. He writes: '[i]t is typically easy for people to accept that colours are illusions and are not really instantiated in the external world, but it is much harder for people to accept that phenomenal properties are illusions and are not really instantiated in our minds'. So, it may be that revelation-c is not as deeply ingrained in our ordinary thinking as revelation-q, which would mean that the alleged intuition of distinctness regarding colour properties is easier to resist than the intuition of dualism.¹⁴

¹⁴ For arguments against the idea that revelation-c is part of our folk theory of colour, see Jackson 1998. It is also interesting to note that the rejection of revelation-c, as opposed to the denial of revelation-q, has a longer history in philosophical discussion. One can find it in the writings of Galileo, Descartes and Locke. Locke's influence, in particular, extends beyond the philosophical circle – his work was widely read in the English-speaking world (Wierzbicka 2006b: 265; Yolton 1977: 8). Discussion in section 6 on Pereboom's proposal for explaining away the intuitiveness of revelation-q also goes some way to supporting the idea that it is easier to conceive the falsity of revelation-c than that of revelation-q. However, a more careful discussion on this issue is nevertheless called for.

Here I am leaving the issue open. It seems that if there are any other types of properties besides phenomenal properties which generate an intuition of distinctness like our intuition of dualism, colour properties seem to be a good candidate. As we saw, revelation-c shares a number of crucial parallel features with revelation-q. This allows us to reach a conditional conclusion, namely, that *if* revelation-c is as deeply ingrained in our ordinary thinking as revelation-q, then there could indeed be a shared, persistent intuition that colour is nonphysical.

5. WHY IS REVELATION INTUITIVE?

We have seen that the thesis of revelation regarding phenomenal properties is intuitive in the sense of being a deep feature of our ordinary conception of experience. We saw two arguments in favour of the latter claim. In chapter 6, we saw that the thesis of revelation being part of our ordinary conception of experience plausibly provides the best explanation for certain linguistic data, i.e. the oddness in sentences such as ‘I know what an itch feels like, but I don’t know what the feeling of an itch really is’. In section 2 of this chapter, borrowing ideas from Peacocke (2008) on implicit conceptions, we explained the intuition of dualism by appealing to our tacit or implicit understanding of the thesis of revelation, which is part of our ordinary conception of experience. This allows us to say that the thesis of revelation being part of our ordinary conception provides a satisfactory (perhaps the best) rational explanation of the intuition of dualism.

The intuitiveness of revelation naturally opens up the question of why revelation is intuitive in the first place, that is, how revelation comes to be part of our ordinary

conception of experience. Furthermore, given the intuitiveness of revelation, there are two potential dialectical contexts in which the revelation argument against physicalism, which we saw in chapter 5, may be contested.

In the first dialectical context, one might think that the intuitiveness of revelation provides *prima facie* support for the truth of the thesis of revelation, which features as a premise in the revelation argument against physicalism. A physicalist might then want to argue that there are alternative explanations for the intuitiveness of revelation which do not require the truth of the thesis of revelation. In this way, the physicalist can try to explain away the intuitiveness of revelation and undermine the support it lends to the thesis of revelation.¹⁵

In the second dialectical context, the physicalist can simply level direct arguments against the thesis of revelation with the aim of undermining the revelation argument against physicalism. Even if the thesis of revelation is false, there must be an explanation as to why it is intuitive, but the physicalist might have no view on what such an explanation might turn out to be. She can accept that the thesis of revelation is part of our ordinary conception and maintain that it is false without having to explain away the intuitiveness of revelation, that is, to supply an explanation for the intuitiveness of revelation without appealing to the truth of revelation.¹⁶

¹⁵ An advocate of revelation, in response, might argue that the truth of the thesis of revelation provides a *better* explanation of the intuitiveness of revelation than the alternative explanations, or put forward more direct arguments in favour of the thesis of revelation.

¹⁶ An advocate of revelation, in this second dialectical context, might respond by arguing that, in fact, the truth of the thesis of revelation is the best explanation for the intuitiveness of revelation, or raise objections against the physicalist's arguments for the falsity of the thesis of revelation.

We will look at objections against the thesis of revelation in the next chapter. In the rest of this chapter, we proceed from the first dialectical context, assuming that the intuitiveness of revelation provides *prima facie* support for the truth of revelation, and we will be concerned with the question of what explains the intuitiveness of revelation.

What then explains the intuitiveness of revelation? One might say that revelation is built into our ordinary conception of experience because it ‘seems obvious’ (Lewis 1995).¹⁷ But then the question of what explains the intuitiveness of revelation becomes the question of what explains why revelation ‘seems obvious’. I will not claim that revelation seems obvious, but I think it is very plausible that we believe or are disposed to judge the thesis of revelation to be true. Indeed, I think if a claim *p* belongs to the ordinary conception of the relevant domain (i.e. if the claim is *intuitive* in this relevant sense), then we would judge or be disposed to judge *p* to be true (i.e. there is an *intuition* that *p*). So plausibly, the *intuitiveness* of revelation provides a natural rational explanation for the *intuition* of revelation.¹⁸ (Note on terminology: by ‘the intuitiveness of revelation’, I here simply mean that revelation is part of the ordinary conception of experience, i.e. our possessing an implicit conception of experience that includes the thesis of revelation; by ‘the intuition of revelation’, I mean the belief/judgement or disposition to believe/judge that revelation is true.)

¹⁷ Lewis (1995) himself in fact thinks that it is because revelation is part of our ordinary conception of experience that it seems obvious. He acknowledges that some philosophers think it is because revelation ‘seems obvious’ that it gets to be part of the ordinary conception of experience.

¹⁸ We can say that the intuitiveness of revelation is explanatorily prior to the intuition of revelation.

Assuming that the intuition of revelation, like the intuitiveness of revelation, also lends *prima facie* support to the thesis of revelation, physicalists would then want to explain away the intuition of revelation – explain it in a way that does not appeal to the truth of revelation. In the rest of this section and the next, I will be concerned with explaining both the *intuition* and the *intuitiveness* of revelation.

In section 6, I will be concerned with a particular physicalist strategy for explaining away the intuitiveness/intuition of revelation.¹⁹ In the rest of the present section, I shall first show how the truth of revelation might feature in an explanation for the intuition of revelation, and then contend that the truth of revelation also plausibly explains the intuitiveness of revelation.

To explain the intuition of revelation, we need to explain why we judge or are disposed to judge the following two claims, which correspond to the two components of the thesis of revelation:

- (i) By having an experience with phenomenal property Q,
one is in a position to know *de dicto* that Q is X.
- (ii) 'X' captures the essence of Q.

If the thesis of revelation is true, then it must be the case that phenomenal property Q *manifests* its essence – which can be captured by the predicate 'X' – to the subject when the subject has an experience with Q. We can think of this manifestation as a kind of

¹⁹ There might also be different styles of explanations for the intuitiveness of revelation: rational, mechanistic, evolutionary, etc. Here, we are particularly interested in whether there could be an explanation for the intuitiveness of revelation that is compatible with the physicalist ontology.

accurate phenomenal appearance of phenomenal properties – in having an experience with Q, it *phenomenally appears* to the subject that Q is X where ‘X’ captures the essence of Q. This accurate phenomenal appearance is not a separate experience that is distinct from the experience with quale Q. If the essence of Q is fully manifested in experience, then a phenomenal appearance of Q as X is just the experience with Q – there is simply no appearance/reality distinction with respect to Q.

This phenomenal appearance that Q is X, given revelation, is crucial in explaining the intuition of (i). Here we shall note that the notion of *phenomenal appearance* is to be distinguished from that of *epistemic appearance*. The distinction corresponds to two uses of perceptual verbs such as ‘seem’, ‘appear’, ‘look’, etc. (see Chisholm 1957; Jackson 1977). According to the epistemic use, the proposition ‘It seems/appears/looks to S that p’ implies that S *believes* that p, e.g. ‘It seems/appears/looks to me that the economy is slowing down’. According to the phenomenal use, the proposition ‘It seems/appears/looks to S that p’ does not necessarily imply that S believes that p; it merely describes some aspects of S’s experience. For instance, in the case of the Müller-Lyer illusion, one might say ‘it seems/appears/looks to me that the two lines are not of the same length’, without implying that one believes that the two lines are of different lengths.

So there is a distinction between a phenomenal appearance, i.e. an experience, that Q is X, and an epistemic appearance, i.e. a belief, that Q is X. But the two ‘appearances’ are often intimately linked. Our beliefs are often reasonable responses to our experiences. We can say that upon having an experience in which it phenomenally

appears that Q is X, the subject is rationally disposed to judge that Q is X and, in normal circumstances, this judgement constitutes knowledge.

Given the aforesaid, we can say that

- (A) our having experiences of which the thesis of revelation is true;

together with

- (B) our possessing the concept of conscious experience in general;
- (C) our possessing some concepts of specific phenomenal properties or kinds of experience; and
- (D) our possessing the concept of knowledge

explain our actually making or being disposed to make particular judgements of the form (i) – By having an experience with phenomenal property Q, one is in a position to know *de dicto* that Q is X.

Now, since experience does not present Q as anything else but X, the addition of

- (E) our possessing the concept of essence (i.e. immediate constitutive essence)

then explains why we would judge or be disposed to judge, not only that 'Q is X', but also that 'Q is X' is an essence-capturing truth about Q, i.e. (ii). This seems plausible especially since, without philosophical theorising, there is no defeater available to the

subject to compel her to reject (ii).²⁰ One might also think that in general we have a tendency to take things at face value and that this tendency is particularly strong in the case of phenomenal properties. This is no surprise if the thesis of revelation is part of our implicit conception of experience. The thesis of revelation entails the claim that there is no (phenomenal) appearance/reality distinction with respect to phenomenal properties.

These particular judgements of the form (i) and (ii) add up to instances of the thesis of revelation. As we just saw, our having experiences of which the thesis of revelation is true, together with our possessing concepts of experience, kinds of experience, knowledge, and essence, explain why we judge or are disposed to judge the thesis of revelation to be true, i.e. the intuition of revelation. We also know, as we saw earlier, that a very natural rational explanation for our intuition of revelation is the intuitiveness of revelation, i.e. our possessing an implicit conception of experience that incorporates the thesis of revelation. Now presumably, having experiences is developmentally prior to having thoughts about experiences and also prior to forming implicit conceptions about experience. I thus conjecture that the thesis of revelation comes to be part of our ordinary conception of experience, i.e. we come to form an implicit conception whose content includes the *general* thesis of revelation, because we have had *particular* experiences of which the general thesis of revelation is true.

Indeed, it seems rather natural to suppose that experiences of instances falling under a general thesis might figure in the aetiology of possession of an implicit

²⁰ As an example of a philosophical-theoretical defeater, if one holds a reductionist worldview and is inclined to hold physicalism, then one would think that phenomenal properties like Q have hidden essences and that 'Q is X' is not an essence-capturing truth about Q.

conception with that general thesis as its content. On this proposal, having experiences of which the thesis of revelation is true not only contributes to explaining why we judge or are disposed to judge particular instances of the thesis of revelation, i.e. the intuition of revelation, but also explains our possession of an implicit conception part of whose content is given by the general thesis of revelation, i.e. the intuitiveness of revelation.

In the next section, we shall be concerned with whether physicalists can successfully explain away the intuitiveness/intuition of revelation. Unlike physicalists' efforts to explain away the *intuition of dualism*,²¹ explicit attempts from the physicalist camp to explain away the *intuitiveness/intuition of revelation* are virtually non-existent. However, Pereboom's discussion (2011) on what he calls the 'qualitative inaccuracy hypothesis' comes close to such an attempt. Although Pereboom does not explicitly talk about revelation or take his hypothesis to entail physicalism, an appeal to this hypothesis can be regarded as a means of explaining away the intuitiveness/intuition of revelation. In the next section, we will assess Pereboom's proposal.

6. THE QUALITATIVE INACCURACY HYPOTHESIS

In his book *Consciousness and the Prospects of Physicalism* (2011), Pereboom (see also 2009) puts forward the qualitative inaccuracy hypothesis (QIH for short):

²¹ Papineau's proposal as we saw in chapter 3 is such an explanation. It acknowledges the intuition of dualism, but argues that it can be explained in a way that does not support the truth of dualism.

[QIH]

It is an open possibility that introspective representation is inaccurate in the respect that it represents phenomenal properties as having qualitative natures they do not in fact have (Pereboom 2011: 14).

QIH says that introspection ‘represents’ phenomenal properties as having ‘qualitative natures’ that they do not have. The two notions, i.e. *qualitative nature* and *introspective representation*, need clarification. Pereboom is not perfectly clear on either of them. Regarding the notion of ‘qualitative nature’, we can take it to mean ‘(intrinsic) features’ of something.²² So QIH says that it is possible that introspection represents phenomenal properties as having features that they do not in fact have. According to QIH, it is possible that such misrepresentation is *systematic*. In this case, there is a systematic discrepancy between what the qualitative nature of a phenomenal property *is* and what introspection represents the qualitative nature *as*. If QIH is true, then phenomenal

²² Pereboom writes:

[W]e introspectively represent phenomenal properties as having specific qualitative natures that are distinct from any *features* that physical theories represent them as having’. (2011: 4, italics added)

Both the physical and introspective modes of presentation represent a phenomenal property as having a specific *qualitative nature*, and the *qualitative nature* that the introspective mode of presentation represents the phenomenal property as having is not included in the *qualitative nature* the physical mode of presentation represents it as having. (2011: 13, italics added)

Pereboom himself does not explicitly endorse the claim that introspection systematically misrepresents the qualitative natures of phenomenal properties. But he thinks that it is an open possibility.

properties might not have any features at all beyond those ascribed to them by, say, physical theories (Pereboom 2011).

Regarding the notion of 'introspective representation', there are at least two possible interpretations here. According to the first interpretation, an introspective representation is an *experience*, i.e. a conscious experience with phenomenal properties. In this sense, introspective representation is understood as 'phenomenal appearance', and is analogous to the notion of 'perceptual representation' or 'perceptual experience'.²³ To say that perception *represents* the car as red is just to say that it *phenomenally appears* that the car is red. Similarly, to say that introspection represents Q as X is just to say that it phenomenally appears that Q has feature X.²⁴ Now, according to QIH, it could be the case that there is systematic introspective representation (i.e. phenomenal appearance) of Q as having the feature X, but Q in fact does not have X. In this case, the introspective representation is an experience that is distinct from the original experience with phenomenal property Q. The original experience *instantiates* (rather than represents) Q and Q *lacks the feature X*. The introspective experience *represents* (in fact, misrepresents)

²³ Note that in section 5, the notion of 'phenomenal appearance' is used, on the assumption that the thesis of revelation is true, in a way that does not admit an appearance/reality distinction with respect to phenomenal properties. In that case, phenomenal appearance is unlike perceptual representation because the latter certainly admits an appearance/reality distinction with respect to what it represents.

²⁴ In this section, 'X' is used to denote a feature of Q that introspection represents Q as having. So if introspection *misrepresents* Q as having X, then X is not a feature that Q has. An advocate of revelation denies a systematic misrepresentation regarding the qualitative nature of Q. She would say that if introspection systematically represents Q as having X and nothing else, then X constitutes the qualitative nature or essence of Q.

the original experience's phenomenal property *Q as having feature X*. In 6.3.1, we will discuss whether QIH is plausible on this interpretation of introspective representation.

There is a second interpretation of the notion of 'introspective representation'. According to this interpretation, an introspective representation is a *belief*. It is understood as 'epistemic appearance', and is analogous to the notion of 'perceptual belief'. We can draw a distinction between perceptual belief and perceptual representation, i.e. perceptual experience. For instance, I have a perceptual representation of the car as red, that is, the car *phenomenally appears* to me to be red; but I might nevertheless have the *perceptual belief* that the car is pink. We can take a particular perceptual belief or judgement regarding object *o* to result from the interplay of my perceptual representation/experience of *o* and my background beliefs about *o*, my viewing conditions, etc. (see Shoemaker 1994: 276-8). If introspective representations are understood as beliefs, then they are analogous to perceptual beliefs in the sense that an introspective belief is based on the interplay of a particular experience and some background beliefs. In this sense, to say that introspection *represents* a phenomenal property *Q as X* is just to say that the subject of introspection introspectively *believes* or *judges* that *Q is X*.

In discussing and motivating his QIH, Pereboom does not make clear which notion of introspective representation he has in mind.²⁵ In the next two subsections,

²⁵ In his 2011 book, Pereboom (2011: 22-3, fn33) seems to suggest that he is thinking of introspective representation as experience or phenomenal appearance. In his replies to critics (2013: 758), he seems to be open to adopt the assumption that introspective representations are beliefs.

which spell out how a physicalist might appeal to QIH to explain the intuitiveness/intuition of revelation (6.1) and how Pereboom motivates QIH (6.2), I will leave the notion of ‘introspective representation’ ambiguous. For the points in these subsections, it does not matter much which interpretation of ‘introspective representation’ is at issue. However, in discussing the plausibility of QIH in 6.3, the question of how precisely we should understand the notion of introspective representation does matter and I will make it clear which interpretation is at issue.

6.1. QIH AND THE INTUITIVENESS OF REVELATION

Given Pereboom’s formulation of QIH, we can see how it might help physicalists explain away the intuitiveness as well as the intuition of revelation. Recall the two claims that make up the thesis of revelation:

- (i) By having an experience with phenomenal property Q,
one is in a position to know *de dicto* that Q is X.
- (ii) ‘X’ captures the essence of Q.

One might take QIH to provide an explanation for the intuition of (i). QIH says that it is possible that introspection systematically represents phenomenal property Q as having a certain qualitative nature, i.e. X, that, in reality, it does not have. An advocate of QIH might say that this introspective representation of Q as X either grounds or else is the subject’s belief that Q is X, disposing the subject to judge (i).²⁶ Regarding the explanation

²⁶ On the assumption that introspective representations are beliefs, to say that introspection represents Q as X is just to say that we introspectively believe that Q is X. On the assumption

for the intuition of (ii), an advocate of QIH can simply adopt an explanation similar to the one suggested in section 5. Since our introspective representation does not represent Q as anything else, it disposes the subject to believe (ii). An advocate of QIH might also say that this systematic introspective misrepresentation of phenomenal properties of experiences as having features that they do not have contributes to explaining why we come to form an implicit conception of experience that incorporates the thesis of revelation, i.e. the intuitiveness of revelation.

6.2. MOTIVATING QIH

Now, if QIH itself is plausible and if QIH supplies a plausible explanation for the intuitiveness/intuition of (i), then physicalists can appeal to QIH to explain the intuitiveness/intuition of revelation, while not themselves endorsing the thesis of revelation. But why should we think that QIH is plausible? Why should we think that there could be a total and systematic misrepresentation of the qualitative natures of phenomenal properties? In Pereboom's writing, one can find two kinds of considerations that he appeals to in motivating QIH. I will clarify these two kinds of considerations in turn, and leave my substantive evaluations to subsection 6.3.

that introspective representations are experiences, the fact that introspection represents Q as X grounds and gives rise to the belief or disposition to believe that Q is X. I shall go on to argue that QIH does not provide a satisfactory explanation for the intuition of (i) in 6.3.

6.2.1. *The Analogy with Colour*

Pereboom considers *an analogy with colour*: QIH is argued to be similar to a qualitative inaccuracy hypothesis in the case of colour properties. On a number of theories of colour, colour properties do not have certain qualitative natures that perception represents them to have.²⁷ For instance, according to both physicalism and dispositionalism about colour, which we saw in section 4, colours, understood as properties of mind-independent objects, do not have the qualitative natures that primitivists take them to have. One might say that such qualitative natures are merely the projections of our own minds. In this sense, physicalists and dispositionalists about colour can be regarded as advocates of a qualitative inaccuracy hypothesis regarding colour properties, which is formulated as below:

[QIH-c]

It is an open possibility that perceptual representation is inaccurate in the respect that it represents colour properties as having qualitative natures they do not in fact have.

Physicalists or dispositionalists about colour would say that our perception *inaccurately* and *systematically* represents C as having a primitive feature Y where C in fact does not have Y. For instance, physicalists about colour might say that the true nature of colour

²⁷ One might think that such a perceptual inaccuracy is an evolutionary adaptation to track similarities and differences in external objects in a relatively simple way (see Chalmers 2018: 25).

property C is the surface reflectance property S, but perception does not accurately represent C *as* S, and instead represents C *as* Y.²⁸

As Pereboom sees it, just as it is possible that perception systematically misrepresents the natures of colour properties, it is possible that introspection systematically misrepresents the natures of phenomenal properties. For Pereboom, since a systematic perceptual inaccuracy regarding the qualitative natures of colour properties as posited by QIH-c is possible, the supposed analogy between QIH and QIH-c then makes it an open possibility that there could be a similar systematic introspective inaccuracy regarding the qualitative natures of phenomenal properties. Of course, this all depends on whether there really is an analogy between QIH and QIH-c, to which I shall return later.

6.2.2. *The Anaesthetic Case*

Pereboom also appeals to intuitive cases to support QIH. In his book, Pereboom gives two examples. Here I will just focus on one of them.²⁹ Pereboom (2011: 23) cites the story of his daughter getting an anaesthetic shot at the dentist's:

Rather than simply showing her the needle in advance and then giving her the injection, the dentist hid the needle from her and told her that he would be dropping bits of cold water into her mouth. She didn't flinch. When I asked her afterward whether the experience was unpleasant, she said that she didn't like the drops of

²⁸ Physicalists/dispositionalists about colour might also appeal to QIH-c to try to explain away the intuitiveness/intuition of revelation-c.

²⁹ Pereboom's other example involves a blindfolded college student in a fraternity initiation who judges that he is in pain when he was merely in contact with an icicle. For considerations against the force of this example, see Kammerer 2018b.

water much. But they didn't hurt. In this case, it may be that the dentist's suggestion, together with his hiding the needle, kept her from introspectively representing the qualitative features of the pain state she was actually in as qualitative features of pain; instead, she misrepresented those features as qualitative features of a sensation of cold.

Let's call this example 'the anaesthetic case'. Pereboom's interpretation of the case is summarised as below:

- (I) The girl's experience, *e*, has the phenomenal property of pain.

The girl *introspectively represents* *e* to have the phenomenal property of a cold sensation.

How this interpretation of the case is supposed to support QIH according to Pereboom requires clarification.

On the assumption that there is no systematic misrepresentation of the qualitative natures of phenomenal properties, we can say that in many cases introspection accurately represents the qualitative natures of phenomenal properties, but there can be local errors where introspective misrepresentations occur. In such cases, we can say that introspection represents a phenomenal property as having a qualitative nature that it does not have. Pereboom takes the anaesthetic case to be such a case. Pereboom's interpretation of the case, i.e. (I), invites an alternative interpretation in slightly different terms. On this alternative we can say that the girl's experience instantiates the phenomenal property of pain, but that she introspectively misrepresents the painfulness of pain as having the qualitative nature of the phenomenal property of a cold sensation. Pereboom (2011: 23) takes the case to supply 'reason to believe that we

are infrequently but nevertheless sometimes can be aware of a discrepancy between the real qualitative nature of a phenomenal property and how it is introspectively represented'. He then takes this kind of local error to make QIH more palatable.

One might question this move from cases of local error to a drastic claim about systematic error. Local errors like the anaesthetic case involve at most misrepresenting a phenomenal property as having a qualitative nature when, in reality, it has a subtly different qualitative nature. Even if we grant that such errors can sometimes happen, it is far from clear that they could be so drastic and systematic as Pereboom envisages – misrepresenting a radically different physical nature as some qualitative phenomenal nature (see also Stoljar 2013a: 749, 'making errors sometimes is quite different from doing it systematically'). But the problem with Pereboom's use of the anaesthetic case to support QIH runs deeper than this. As we shall see in 6.3.1, if introspection representation is understood as *experience*, then Pereboom's interpretation of the anaesthetic case is hardly plausible.

6.3. AGAINST QIH

In this subsection, I evaluate QIH (repeated here for convenience):

It is an open possibility that introspective representation is inaccurate in the respect that it represents phenomenal properties as having qualitative natures they do not in fact have (Pereboom 2011: 14).

Earlier, we saw that introspective representations can be understood as either *experiences* or *beliefs*. Given these two interpretations, I shall argue that those who want to appeal to QIH to explain away the intuitiveness/intuition of revelation face a dilemma. According to the first horn of this dilemma, if introspective representations are experiences, then QIH is unmotivated and deeply implausible. According to the second horn, if introspective representations are beliefs, then QIH cannot explain away the intuitiveness/intuition of revelation. Overall, I shall argue that there is no version of QIH that is simultaneously plausible and can help physicalists satisfactorily explain away the intuitiveness/intuition of revelation.

6.3.1. First Horn: Introspective Representations as Experiences

On the experience-conception of introspective representation, QIH is understood as below:

[QIH-ph]

It is an open possibility that introspective representation (understood as phenomenal appearance or introspective experience) is inaccurate in the respect that it represents phenomenal properties, such as *Q*, as having features they do not in fact have.

In this subsection, I shall argue that on the experience-conception of introspective representation, not only do the analogy with colour and the anaesthetic case fail to motivate QIH-ph, the hypothesis itself is also deeply implausible.

Let's consider Pereboom's anaesthetic case first. On the interpretation that introspective representation is *experience* or *phenomenal appearance*, Pereboom's interpretation of the case is to be understood as follows:

(Ia) The girl's experience, *e*, has the phenomenal property of pain.

The girl *introspectively represents* *e* to have the phenomenal property of a cold sensation. That is, it *phenomenally appears* to the girl that *e* has the phenomenal property of a cold sensation.³⁰

The girl introspectively believes *e* to have the phenomenal property of a cold sensation.

In this case, there is not only a false introspective belief, but also a false introspective appearance, i.e. an introspective misrepresentation.

Now, consider alternative interpretations of the anaesthetic case:

(Ib) Experience *e* has the phenomenal property of pain.

Experience *e* *phenomenally appears* to the girl to have the phenomenal property of pain.

The girl introspectively believes *e* to have the phenomenal property of a cold sensation.

³⁰ The notion of phenomenal appearance at issue is about what *phenomenal property* the experience phenomenally appears to the subject as having, not about what *qualitative nature* the phenomenal property phenomenally appears to the subject as having. The latter kind of phenomenal appearance features in QIH-ph. Here we are not dealing with QIH-ph, but are only concerned with how best to interpret the anaesthetic case and whether it supports QIH-ph.

(Ic) Experience e has the phenomenal property of a cold sensation.

Experience e *phenomenally appears* to the girl to have the phenomenal property of a cold sensation.

The girl introspectively believes e to have the phenomenal property of a cold sensation.

Neither (Ib) nor (Ic) involves any introspective misrepresentation understood as phenomenal appearance, i.e. experience, and would thus in no way support QIH-ph. On Pereboom's interpretation, i.e. (Ia), there is a distinction between what phenomenal property the girl's experience phenomenally appears to her as having and what phenomenal property her experience in fact has. This dubious *appearance/reality distinction* regarding the reality of experience, which is absent in (Ib) and (Ic), is precisely, as I shall show, what makes (Ia) worse than its alternatives as an interpretation of the anaesthetic case.

In the literature on introspection, in discussing whether there is an appearance/reality distinction with respect to phenomenal experiences and their phenomenal properties, philosophers distinguish between two kinds of appearance/reality distinction, corresponding to the distinction between phenomenal appearance and epistemic appearance, which we saw in section 5.³¹ Some philosophers have argued for an '*epistemic appearance/reality distinction*': we can make errors in introspective judgements about what phenomenal properties our experiences have (see

³¹ Note that the relevant appearance/reality distinctions in the literature do not pertain to phenomenal properties and their natures.

Schwitzgebel 2008, 2011). In contrast, no one has seriously proposed a ‘*phenomenal* appearance/reality distinction’ – there is no distinction between what phenomenal property the experience *phenomenally appears* to the subject to have, and what phenomenal property the experience in fact has. The idea that there is no distinction between the phenomenal appearance of an experience and the reality of an experience is widely regarded as plausible (see Moran 2001: 14; Gertler 2012: 127; Searle 1997: 456; Horgan 2012: 406).

If introspective representations are experiences, i.e. phenomenal appearances, then Pereboom’s interpretation (Ia) of the anaesthetic case is committed to an implausible phenomenal appearance/reality distinction. The latter distinction is absent in interpretations (Ib) and (Ic). Interpretation (Ib) merely posits an epistemic appearance/reality distinction, whereas (Ic) posits no appearance/reality distinction of any sort. Since Pereboom’s interpretation of the anaesthetic case seems to fare worse than alternative interpretations which in no way support QIH-ph, the anaesthetic case fails to motivate QIH-ph.

On the experience-conception of introspective representation, QIH-ph fails to be motivated by an analogy with colour. The analogy between QIH and QIH-c, as we saw in 6.2.1, relied on an alleged analogy between introspective representation and perceptual representation, i.e. perceptual experience. Those who maintain that there is a relevant similarity between the two (i.e. between introspection and perception, where perception is conceived along the lines proposed by representationalists) hold the inner sense theory of introspection, or what Shoemaker (1994) calls ‘the broad perceptual

model' of introspection (see also Armstrong 1968; Gertler 2017). On this theory, there is a similarity between introspective representations and perceptual representations in the sense that the relationship between introspective representation and what is introspectively represented, i.e. a mental state, can be thought of as *causal* and *independent*, just like the relationship between perceptual representation and what is perceptually represented, i.e. external objects and their properties (Shoemaker 1994: 271). Setting aside the issue of whether the theory itself is plausible (see Shoemaker 1994), it seems that on this theory, insofar as introspective representation is thought to be similar to perceptual representation, the notion of introspective representation is understood as belief, not experience (see Shoemaker 1997: 271). As Richard Moran (2001: 14) points out, no inner sense theorist would say that one's introspective belief that one is having a headache is 'mediated by an appearance of the headache'; on this theory, 'there is simply nothing quasi-experiential in the offing to begin with'. So, if there is no plausible analogy between perceptual representation and introspective representation where both the former and the latter are conceived as experience, then there is no analogy between QIH-c and QIH-ph. Consequently, QIH-ph becomes unmotivated.

This leads to my argument that QIH-ph is deeply implausible if introspective representation is understood as experience. Suppose I have an experience e_1 with phenomenal property Q which has a physical nature N. On QIH-ph, where introspection systematically misrepresents Q as X, the introspective representation/experience is a

further experience (i.e. phenomenal appearance), e_2 , that represents, and occurs simultaneously with, but is distinct from e_1 .³²

If the idea that there is a separate experience e_2 mediating between experience e_1 (which e_2 misrepresents) and introspective beliefs about e_1 , does not make QIH-ph bizarre enough, then the following consideration should bring home my claim that QIH-ph is implausible. Now, since introspective representation e_2 is an experience, it must, given that experiences have phenomenal properties, also have some phenomenal property – call this property Q' – which presumably has a physical nature N' . But Q' does not seem to have a physical nature – its nature certainly seems to be more like X than N . Given this consequence, an advocate of QIH-ph is forced to say that there is also an introspective misrepresentation regarding the nature of Q' – Q' is misrepresented as having a qualitative nature (X' , say) that it does not in fact have. But if introspective misrepresentation applies to both first-order and high-order phenomenal properties – that is, if all phenomenal properties are such that introspection systematically misrepresents their natures – then we are left with a bizarre claim that there is an infinite regress of introspective experiences. First, we have a second-order experience, e_2 , of the first-order experience e_1 with quale Q . Given that e_2 also has a phenomenal property, there is a third-order experience, e_3 , of the second-order experience e_2 with quale Q' . Since e_3 is also an experience and presumably has a phenomenal property, there would have to be a simultaneous fourth-order experience. The pattern iterates. This infinite

³² Indeed, the latter claim follows if introspective representation/experience is understood on the model of perceptual representation/experience as the analogy between QIH-ph and QIH-c requires.

regress of simultaneous but distinct introspective (mis)representational experiences is a profoundly implausible commitment of QIH interpreted as QIH-ph.

Since on the experience-conception of introspective representation, QIH is deeply implausible and lacks motivation, if physicalists are to appeal to QIH to explain the intuitiveness/intuition of revelation, then they had better understand the relevant notion of introspective representation as belief.

6.3.2. *Second Horn: Introspective Representations as Beliefs*

On the belief-conception of introspective representation, QIH is understood as below:

[QIH-ep]

It is an open possibility that introspective representation (understood as epistemic appearance or introspective belief) is inaccurate in the respect that it represents phenomenal properties, such as Q, as having features they do not in fact have.

In this case, there is *no* separate phenomenal appearance (experience e_2) misrepresenting Q as X and mediating between an experience e_1 with quale Q and the introspective judgement or belief that Q is X. The introspective misrepresentation of Q as X is inaccurate in the sense that the judgement that Q is X is erroneous, because, as Pereboom would say, Q does not have X as a feature.

This reading of introspective representation makes Pereboom's interpretation of the anaesthetic case plausible:

(Ib) Experience e has the phenomenal property of pain.

Experience e phenomenally appears to the girl to have the phenomenal property of pain.

The girl *introspectively represents*, that is, introspectively *believes*, e to have the phenomenal property of a cold sensation.

We can say that the introspective misrepresentation, i.e. a misjudgement resulting in a false *belief*, occurs because the girl did not pay attention to the phenomenal character of her experience, and her judgement was influenced by her expectations, e.g. that the dentist was going to drop cold water in her mouth.

But on the belief-conception of introspective representation, there is no sufficient analogy between QIH-ep and QIH-c since QIH-c is formulated in terms of phenomenal representation which is understood as experience not belief. So there seems to be no suitable analogy with colour that would motivate QIH-ep.

Furthermore, as I shall argue, QIH-ep does not allow the physicalist to explain away the intuitiveness/intuition of revelation. Recall the two components of the thesis of revelation:

- (i) By having an experience with phenomenal property Q,
one is in a position to know *de dicto* that Q is X.
- (ii) 'X' captures the essence of Q.

Explaining the intuition of (i) is precisely what is crucial in explaining the intuition of revelation. A physicalist would have to explain the intuition of (i) without conceding

that Q in fact has feature X. But note that in order to explain the intuition of (i), that by having an experience with quale Q, one is in a position to know – or at least rationally judge – that Q is X, one needs to explain, not only the intuition of:

(i-a) By having an experience with phenomenal property Q, one is in
a position to *believe* or *judge* that Q is X.

but also the intuition of:

(i-b) The belief Q is X is *justified* or *rational*.

I shall argue that QIH-ep fails to provide a satisfactory explanation for the intuition of (i).

To bring out my argument, consider Christopher Hill's distinction between *errors of judgement* and *errors of ignorance*, a distinction Pereboom himself notes in his book (2011: 22-3, fn33):

[I]t is necessary to distinguish between two kinds of errors that we are prone to make. First, as we have just noticed, errors can occur when beliefs are based on appearances that fail to do justice to the entities to which the beliefs refer. When we are misled in this way by imperfect information, we make errors that may be called *errors of ignorance*. Second, errors can arise when we have adequate information about the entities with which we are concerned but we fail to take this information fully into account in forming beliefs about the entities. Errors of this sort, which may be called *errors of judgement*, are usually due either to some form of inattention or to the influence of expectation upon judgement. They tend to occur when we are hasty in forming beliefs, when we are suffering from information overload, when we are preoccupied, and when we are being lazy. They also tend to occur in situations in which anticipation causes us to lower the thresholds of application that are associated with some of our concepts. (Hill 1991: 128)

Errors of ignorance result from misleading phenomenal appearances of things. I erroneously make the perceptual judgement that a car is pink because it phenomenally appears to be pink, but is in fact red. Errors of judgement usually result from certain epistemic shortcomings on the part of the subject. I erroneously make the perceptual judgement that the car is pink despite the fact that it phenomenally appears to me to be red. In this case, the error in my judgement might result from the fact that I failed to attend to my perceptual experience closely. The anaesthetic case, understood as (Ib), is an example of an error of judgement, perhaps resulting from inattention or the influence of the girl's expectations.³³

Now, on the belief-conception of introspective representation, for my false judgement (false by Pereboom's lights) that Q is X to be rational or justified, it would have to be an instance of an error of ignorance based on imperfect information. But on this proposal, there is simply no mediating experience in which Q phenomenally appears to be X. So, appealing to QIH-ep cannot explain the intuition of (i-b). Furthermore, it also fails to explain the intuition of (i-a). My false judgement is not due to 'inattention or to the influence of expectation upon judgement'. No amount of attention could make a difference to my introspective judgement about what qualitative nature Q has since Q's real nature is systematically hidden. My false judgement Q is X also does not seem to be due to the influence of expectation, for I cannot possibly expect Q to be X since I have never had an experience with a quale which has X as its feature.

³³ Regarding the other potential interpretations of the anaesthetic case, (Ia) posits an error of ignorance; and (Ic) posits no error of any sort.

Overall, it seems that, on Pereboom's account, there is nothing that explains the error in my judgement. My false judgement simply arises out of nowhere. But this makes my judgement seem utterly irrational – indeed, difficult to make sense of. So, the appeal to QIH-ep wouldn't give physicalists a satisfactory explanation for the intuition of (i). Not only can QIH-ep not help physicalists explain away the intuition of revelation, it is also utterly unclear how the belief-conception of introspective representation posited by QIH-ep could help explain the intuitiveness of revelation.

In his book, Pereboom (2011: 22-3, fn33) takes the open possibility posited by QIH to 'have us making errors of ignorance in our introspection-based beliefs about phenomenal properties, since such beliefs would be based on *appearances* that fail to do justice to the real qualitative nature of those properties' (italics added). There Pereboom seems to understand introspective representations as experiences. If we allow such mediating experiences, then the false judgement 'Q is X' can be treated as an error of ignorance and in this way we can explain the rationality in our false judgement on the basis of a false appearance. However, as we saw in 6.3.1, this way of understanding introspective representation renders QIH implausible and devoid of motivation.

I conclude that there is no version of Pereboom's QIH such that (1) it is plausible and (2) it can help physicalists, who reject the thesis of revelation, to satisfactorily explain away the intuitiveness/intuition of the thesis. Other proposals to explain away the intuitiveness/intuition of revelation may be available to physicalists. But it remains unclear what such a proposal would look like and whether it would be successful. So, it

remains unclear that physicalists can explain away the intuitiveness/intuition of revelation.

7. CHAPTER CONCLUSION

In this chapter, I argued that the intuition of dualism can plausibly be explained by the fact that the thesis of revelation regarding qualia is intuitive in the sense of being part of our ordinary conception of experience, and I defended the proposed explanation against a number of objections.

In addition, I looked at another area where a similar intuition of distinctness, like the intuition of dualism, could arise. We saw that there is a similar thesis of revelation regarding colour properties, i.e. revelation-c. This thesis of revelation and the thesis of revelation regarding phenomenal properties, i.e. revelation-q, share a number of parallel features. Given these shared features, one might expect an intuition of distinctness regarding colour properties and physical properties. I left it an open question whether colours generate an intuition of distinctness like the intuition of dualism generated by qualia. I concluded that *if* revelation-c is as deeply ingrained in our ordinary thinking as revelation-q, then there could indeed be a shared, persistent intuition that colour is nonphysical.

Finally, I delved into the issue of what explains the intuitiveness of the thesis of revelation (regarding phenomenal properties). I laid out a proposal for explaining the intuitiveness of revelation which draws on the truth of revelation itself. Physicalists may be particularly interested in explaining the intuitiveness of revelation in a way that is

compatible with the physicalist ontology. As we saw, Pereboom's hypothesis about introspective inaccuracy offers such an explanation. However, as I argued, the hypothesis is either implausible, or fails to explain both the intuition of revelation and its intuitiveness. It remains unclear whether there is a satisfactory explanation for the intuitiveness of revelation that is compatible with physicalism. This concludes Part II, and completes the main task set out at the beginning of the dissertation, that is, to address the question of what underpins our intuition of dualism.

Now, while it may be in physicalists' interest to explain away the intuitiveness of the thesis of revelation, the intuitiveness of revelation on its own does not necessarily pose a problem for physicalism. Physicalists can accept my proposed explanation of the intuition of dualism in terms of the intuitiveness of revelation. They can accept, as some physicalists indeed do, that revelation is part of our ordinary conception of experience. But they do need to deny that the thesis of revelation is in fact true. This leads us to Part III of the dissertation. In the next chapter, I will critically discuss various objections one might raise regarding the thesis of revelation as well as regarding the revelation argument against physicalism.

Chapter 8

REVELATION AND OBJECTIONS

1. INTRODUCTION

In previous chapters, we saw that the thesis of revelation is part of the ordinary conception of experience and that its intuitiveness provides a plausible explanation for the intuition of dualism. In this chapter, we consider whether the intuitive thesis of revelation withstands philosophical scrutiny. Direct arguments against revelation are few in the literature, with the notable exception of Stoljar's (2018) review of Goff's (2017) recent book *Consciousness and Fundamental Reality*. Goff (2017) appeals to revelation to argue against physicalism, as we shall see shortly. In his review of Goff's book, Stoljar (2018) raises three objections against Goff's argument: (i) the argument apparently 'proves too little' – it does not show that grounding physicalism is false; (ii) the argument might 'prove too much' – if it shows that grounding physicalism is false, it would also show Russellian monism to be false; and (iii) the thesis of revelation (at least given Goff's formulation) is implausible. The direct target of Stoljar's three objections is *the revelation argument* against physicalism as found in Goff (2017). In chapter 5, I also spelt out the incompatibility between revelation and physicalism. Stoljar's three objections equally target that formulation of the revelation argument. Thus, it is important to respond to all three of Stoljar's objections.

Stoljar's objection (iii), however, is of particular interest here. It is an objection against the *thesis of revelation* itself – it raises an objection against *the revelation argument*

by problematising the *thesis of revelation* which serves as a premise in the argument. Doubt has also been cast on revelation by Kelly Trogon (2017). If revelation is implausible to start with, then the revelation argument is undermined.

In addition to addressing worries raised by Trogon and Stoljar regarding the revelation argument, I also turn to the discussion on introspection in relation to the thesis of revelation. In the literature, the thesis of revelation is often taken to be the claim that *introspection* reveals the essences of qualia (Goff 2015, 2017; Chalmers 2016, 2018; Elpidorou 2016; Majeed 2017). The topic of introspection, however, has generated much discussion recently (e.g. Smithies and Stoljar 2012; Bayne and Spener 2010; Schwitzgebel 2011; Kriegel 2013). In particular, the reliability of introspection as a means of acquiring knowledge about one's own occurrent mental states has been called into question (e.g. Schwitzgebel 2011). One might think that while introspection is reliable at delivering certain kinds of introspective beliefs, its domain of proper functioning is limited (Bayne and Spener 2010). I shall clarify how the limits of introspection pose no problem for the claim that introspection reveals the essences of qualia.

The structure of the chapter is as follows. Section 2 addresses the worry about revelation raised by Trogon (2017). Sections 3-5 respond to Stoljar's three objections against the revelation argument. Section 6 turns to issues to do with the limits of introspection. Section 7 concludes the chapter. Overall, I shall defend the position that while the thesis of revelation is incompatible with various metaphysical theories of consciousness, there is no serious objection against the thesis itself.

2. PARTIAL REVELATION

Revelation, as I have formulated it in chapter 4, is the claim that by having an experience with phenomenal property Q, one is in a position to know that Q is X where 'X' captures the (full immediate constitutive) essence of Q. In a recent paper, Trogon (2017) may be taken as claiming that what seems plausible is in fact the thesis according to which what one knows in having an experience with phenomenal property Q is only *part of the essence* of Q – call this 'partial revelation'.¹

In the paper, Trogon (2017: 2347) asks us to consider the phenomenal property of *phenomenal red*, which, according to him, has *phenomenal red** – 'the property *being such that if a person has a phenomenally red state then the person feels like thus-and-so in virtue of this fact*' – essentially, i.e. *phenomenal red** is an essential property of *phenomenal red*. According to Trogon (2017: 2350), the fact that *phenomenal red* has *phenomenal red** as an essential property is compatible with *phenomenal red* having 'some other property essentially as well'. To provide an example, he (2017: 2350) says 'suppose that *phenomenal red* is a physical/functional property. It follows that *phenomenal red* has *being a physical/functional property* essentially'. But experience does not, Trogon would say, reveal this latter property, i.e. *being a physical/functional property*, as belonging to the essence of *phenomenal red*, i.e. experience does not reveal that *phenomenal red* is a physical/functional property.

¹ In his paper, Trogon (2017) formulates this thesis in terms of the notion of *phenomenal concepts* and what he calls 'essential characterisation'. The details of Trogon's system should not matter here.

In formulating the thesis of revelation in chapter 4, I appealed to Fine's (1995a, 1995b) notion of *immediate constitutive* essence – what a thing is in its most core respects. Although he does not make this explicit, I don't think Trogon understands the notion of essence as Fine's *immediate constitutive* essence. Trogon (2017: 2350, 'It follows that ...') appears to argue that for any physical property P, the property *being a physical property* belongs to the essence of P, as he uses the term 'essence'. Now, consider the property of *being a sister*. Everyone would agree that *being a sister* is a physical property, but *being a physical property* does not seem to belong to the *immediate constitutive* essence of *being a sister*. It seems that what it is to be a sister in *its most core respects* is to be a female sibling.

Consider also Trogon's example of *phenomenal red*. Trogon says that a subject feels like *thus-and-so* in virtue of being in a phenomenally red state. In other words, the way the subject feels in virtue of being in a phenomenally red state is *thus-and-so*. That is, *what it is like for the subject* to be in a phenomenally red state, i.e. *phenomenal red*, is *thus-and-so*. Once *phenomenal red* is understood this way, it seems intuitive that there is no more to the *immediate constitutive essence* of *phenomenal red* than its '*what it is like for the subject*' property, its '*thus-and-so-ness*' (which Trogon aims to capture with '*phenomenal red**'). Neither *being physical/functional* nor *being an event of R-fibres firing* seems to be part of what the property *phenomenal red* is in its most core respects.

Note that the thesis of revelation, as formulated in chapter 4, says that the subject of an experience e with quale Q is in a position to know *that Q is X*, where (in fact) 'X' captures the (full) immediate constitutive essence of Q. On this view, even if *phenomenal red* were identical to *being an event of R-fibres firing*, it would not follow that the immediate

constitutive essence of *phenomenal red* would include *being physical/functional*. It would also not follow from the thesis of revelation that an experience with *phenomenal red* would reveal to the subject of the experience that *phenomenal red is physical/functional*. So, it would not be inconsistent with the thesis of revelation that an experience with phenomenal property Q does not reveal that Q is physical/functional.

Given that Trogon apparently understands the notion of essence in a broad sense that goes beyond the notion of immediate constitutive essence, his discussion does not raise any objection against the thesis of revelation, since revelation, as I have formulated it, is a claim about knowing the *immediate constitutive* essences of qualia. My formulation of revelation nevertheless provides a starting point for an argument against physicalism about phenomenal properties (as we saw in chapter 5). If *phenomenal red* were *being an event of R-fibres firing*, then that would plausibly be the essence of *phenomenal red*. In that case, it would be inconsistent with the thesis of revelation that experience does not reveal *phenomenal red* to be *R-fibres firing*.² Thus, it follows from the thesis of revelation that *phenomenal red* is not *being an event of R-fibres firing* – nor any other physical/functional property.

² Here it is worth pointing out that, even on Trogon's broad notion of essence, it is not clear that his 'partial revelation' helps physicalism. On his 'partial revelation' and on his notion of essence, a phenomenal property Q might have a number of 'essential properties', and in having an experience with Q, one is only in a position to know a subset of these properties but never the whole set. Given this view, even if phenomenal red were to be a physical/functional property, e.g. being an event of R-fibres firing, it is still supposed to have a 'what-it-is-like' property or feature (which Trogon aims to capture as *phenomenal red**). It is not clear that *that* property or feature would be physical/functional.

In the rest of the chapter, I shall continue to use the Finean notion of immediate constitutive essence and I shall use ‘revelation’ to refer only to the thesis formulated in chapter 4. Let’s now turn to Stoljar’s three arguments against *the revelation argument*, that is, the argument against physicalism which features revelation as a crucial premise.

3. ‘PROVES TOO LITTLE’

According to Stoljar, the revelation argument ‘apparently proves too little’. Here Stoljar has in mind Goff’s formulation of the revelation argument, which Goff succinctly puts as follows:

[W]e know what pain is through feeling pain, and hence if pain were c-fibers firing, we'd know about it. But we don't, so it isn't. (Goff 2017: 125)

By ‘we know what pain is’, we can take Goff in this context to mean that we know what the essence of quale pain is (i.e. we know the essence of the phenomenal type that is the experience of pain); and by ‘pain is C-fibres firing’, we can take Goff to mean that quale-pain is identical to the physical property of being an event of C-fibres firing (i.e. the type of conscious experience that is the experience of pain is identical to the type of physical event that is C-fibres firing).³ Stoljar’s contention is that Goff’s revelation argument, as stated above, only shows that *identity physicalism* is false, e.g. that quale-pain is not identical to the physical property of being an event of C-fibres firing; it does not show that other versions of physicalism are false. For instance, one might, as Goff (2017)

³ I think this is a charitable reading of Goff (see also Goff 2017: 107, 2015: 214).

indeed does, formulate physicalism as a grounding thesis, e.g. the instantiation of quale pain is *grounded* in the instantiation of the physical property of being an event of C-fibres firing, rather than *identical* to it. Stoljar thinks that ‘it is not easy to see’ how the argument from revelation can be adjusted to show that *grounding physicalism* is false. So, it seems that Goff’s revelation argument only shows that identity physicalism is false. But this proves too little – there is no serious threat from revelation against physicalism because physicalists need not be identity physicalists. (Note that whether or not the revelation argument proves too little is not an objection against *revelation* itself. It is an objection against *the revelation argument* which is supposed to show that physicalism is false.)

In chapter 5, we saw that revelation is incompatible with various versions of physicalism, including identity physicalism, realiser functionalism, scientific role functionalism and commonsense role functionalism. The former two versions of physicalism take phenomenal properties to have physical essences and identify phenomenal properties with physical properties. The latter two versions of physicalism take phenomenal properties to have functional essences and identify phenomenal properties with functional properties. So revelation is not just incompatible with identity physicalism.

However, in chapter 5, nothing was said about whether revelation is incompatible with grounding physicalism. In the rest of this section and in responding to Stoljar’s first objection against the revelation argument, I shall argue that the revelation argument does apply to grounding physicalism given a certain formulation of the latter. In spelling out the tension between grounding physicalism and the thesis

of revelation, I will draw on Dasgupta's (2014) well-known formulation of grounding physicalism.

Before we turn to grounding physicalism, let's first clarify the notion of grounding. Grounding is commonly understood as a non-causal explanatory relation that holds between facts.⁴ Consider the example involving *conferences* which Dasgupta (2014: 566) appeals to in illustrating the notion of grounding:

- (F) The fact that an event *e* contains people engaged in C-activities
 (i.e. giving talks, listening to talks, asking questions, and so on)
 grounds the fact that *e* is a conference.

Event *e* is a conference *because* (in the non-causal sense of the word) there are people engaged in C-activities, that is, giving talks, listening to talks, asking questions, etc.

Facts like (F) are called 'grounding facts' – they are facts about grounding. One might ask: What grounds grounding facts? Why does the event *e* containing people engaged in C-activities make the event *e* a conference? A natural answer, as Dasgupta (2014: 567) puts it, is that a conference is just 'the kind of thing that you get when people engage in those activities'. So there is a *general connection* between conferences and C-activities, which grounds the grounding fact (F). Dasgupta goes on to argue that such a general connection is groundless.

⁴ In the context of discussing the nature of phenomenal properties, it is also common to talk about grounding as a relation holding between properties. One might then have the locution: 'phenomenal properties are grounded in physical properties' (see Chalmers 2016, Stoljar 2018).

There are different ways to formulate this idea of a general connection, which give rise to different versions of what Dasgupta (2014: 568) calls ‘brute connectivism’: it could be an *essential* truth (‘brute essentialism’), or a *necessary* truth (‘brute necessitarianism’), or a *conceptual* truth (‘brute conceptualism’), or a *metaphysical law* (‘brute nomicism’), such that if an event includes people engaged in C-activities then it is a conference. Dasgupta himself focuses on *brute essentialism*. On this view, conferences are *essentially* events consisting in people engaged in C-activities. Given brute essentialism, (F) is understood as being grounded in the following two facts (Dasgupta 2014: 568):

(F.i) Event *e* contains people engaged in C-activities.

(F.ii) It is essential to being a conference that if an event contains people engaged in C-activities then it is a conference.

Facts like (F.ii) are essential facts – facts about the essence of a thing.⁵ Dasgupta goes on to argue that essential facts are groundless and in particular, they are ‘autonomous’ – they are ungrounded and are also not apt for being grounded.⁶

Now, grounding physicalists claim that phenomenal facts are ultimately grounded in physical facts at the fundamental level. For instance, they might claim:

⁵ In his paper, Dasgupta (2014: 589) makes it explicit that the notion of essence he is operating with is Fine’s ‘immediate constitutive essence’. To say that it is *essential* to A that B, means that B is part of the immediate constitutive essence of A.

⁶ *Autonomous* facts are contrasted with what Dasgupta calls ‘substantive facts’ which are apt for being grounded. With respect to a substantive fact, ‘the question of what grounds it can legitimately be raised and given a sensible answer, an answer that either states its ground or else states that it has none’ (Dasgupta 2014: 575). In contrast, ‘the question of what grounds it does not legitimately arise’ with respect to an autonomous fact (Dasgupta 2014: 576).

(G) The fact that S is undergoing an event with physical property P
grounds the fact that S is undergoing an experience with
phenomenal property Q.

(G) is a grounding fact. As with (F) regarding Dasgupta's example of a conference, a grounding physicalist can appeal to *brute essentialism* to spell out the ground for (G). I shall discuss two ways of spelling out the ground for (G), and show that neither version of grounding physicalism is compatible with revelation.

A grounding physicalist might say that (G) is grounded in the following two further facts:

(G.i) S is undergoing an event with physical property P.

(G.ii) It is essential to phenomenal property Q that if something is undergoing an event with physical property P, then this something is undergoing an experience with phenomenal property Q.⁷

(G) obtains *because* (G.i) and (G.ii). (G.ii) posits a tight connection between the essence of the phenomenal and the physical. According to (G.ii), knowing the essence of Q requires

⁷ The above way of formulating (G.ii) follows Dasgupta (2014). In his paper, Dasgupta (2014: 560) focuses on the following grounding fact:

(C) The fact that subject SD's brain is in physical state P grounds the fact that SD is conscious.

Dasgupta (2014: 580) takes (C) to be grounded in the following two facts:

(C.i) SD's brain is in physical state P.

(C.ii) It is essential to being conscious that if something's brain is in physical state P then it is conscious.

one to know about physical property P and know *that* if P is instantiated then Q is instantiated. It is straightforward that (G.ii) is incompatible with revelation. It is not true that when I have an experience with quale Q, I am in a position to know that Q is such that if I am undergoing an event with physical property P, then I am undergoing an experience with Q. So a grounding physicalist who appeals to (G.ii) to ground (G) would have to reject revelation.

One might contend that (G.ii) is too strong. Consider role functionalism, which takes phenomenal properties to have functional essences. If role functionalism is true, then Q has a functional essence that is multiply realised by different physical properties playing the same functional role. A grounding physicalist who is also a role functionalist would agree to (G). But she need not commit to (G.ii), that it is part of the essence of Q – where Q is a functional property – that if a particular physical realiser P is instantiated Q is instantiated. On the role functionalist's picture, one can know the essence of Q, that is, know its functional role, without knowing which physical realiser properties play that role. So, one might plausibly claim that to know the essence of Q does not require one to know anything about the ground; nevertheless, it is part of the essence of Q that it is grounded in some property or other, where, a grounding physicalist would claim, this property turns out to be physical.

Consider the following weaker alternative to (G.ii) that a grounding physicalist might propose:

(G.ii*) It is essential to phenomenal property Q that it is grounded in
some property such that if something is undergoing an event

with this property, then this something is undergoing an experience with phenomenal property Q.⁸

In this case, it is essential to Q that it is grounded (though one might know the essence of Q without knowing what the ground is or that the ground is physical). A grounding physicalist would say that, in fact (though not as a matter of Q's essence), the ground for Q in our world is physical, and the particular physical property P can be such a ground.

But (G.ii*) is also incompatible with revelation. It is not true that by having an experience-token with quale Q, I am in a position to know that Q is grounded in some property which is such that if this property is instantiated then Q is instantiated. Nothing in my experience with quale Q informs me of the fact that Q has a ground. So a grounding physicalist who appeals to (G.ii*) to ground (G) would also have to reject revelation.

To recap, the thesis of revelation makes a claim about what the essence of a quale is – the essence of a quale is something we are in a position to know by having an experience with that quale. Grounding physicalism which appeals to the idea of brute essentialism commits to a particular claim about what belongs to the essence of a quale, formulated as either (G.ii) or (G.ii*). But since the latter two claims are incompatible with revelation, such a physicalist would have to deny revelation.

⁸ In addition to being essentially grounded in some property, Q would presumably have other essential properties which distinguish Q from other qualia. For role functionalists, for example, what distinguishes Q from other phenomenal properties would be the specific functional role associated with Q. A role functionalist can say that it is essential to Q that it is grounded in some property and this property is such that it plays the functional role R.

It is worth noting that if one developed grounding physicalism by appealing to other versions of brute connectivism, e.g. brute necessitarianism, brute conceptualism and brute nomicism, it would be less straightforward to demonstrate that the resulting version of grounding physicalism is incompatible with revelation. This is because these other ways of developing grounding physicalism do not make direct claims about what the essence of a quale is; whereas the thesis of revelation does make such a claim. One might think that these other versions of brute connectivism still make *indirect* claims about the essences of qualia, and that the revelation argument also applies to grounding physicalism formulated with these versions of brute connectivism. Given the limited space here, I will not delve into this issue. But as we have seen, if one develops grounding physicalism in terms of brute essentialism, as Dasgupta indeed does, then in response to Stoljar, we have a straightforward argument from the thesis of revelation against such a version of grounding physicalism. The argument is summarised below:

- (1) If S has an experience-token with phenomenal property Q, then S is in a position to know that 'Q is X', where the predicate 'X' captures the essence of Q.
- (2) If grounding physicalism is true, then it is essential to Q that Q is grounded in some property which is such that if this property is instantiated then Q is instantiated.⁹

⁹ Here I am operating with the weaker principle (G.ii*).

- (3) If it is essential to Q that Q is grounded in some property which is such that if this property is instantiated then Q is instantiated, then by having an experience-token with phenomenal property Q, S should be in a position to know that 'Q is grounded in some property which is such that if this property is instantiated then Q is instantiated'.
- (4) It is not true that by having an experience-token with phenomenal property Q, S is in a position to know that 'Q is grounded in some property which is such that if this property is instantiated then Q is instantiated'.
- (5) Grounding physicalism is false.

Premise (1) is entailed by the thesis of revelation (see chapter 4, section 4). (2) states what is essential to a quale if grounding physicalism were true. (3) follows from (1) and (2). (4) seems intuitive. (5) then concludes from (3) and (4) that grounding physicalism is false.

Given the discussion in this section, we can conclude that, *contra* Stoljar (2018), it is far from clear that the revelation argument proves too little.

4. 'PROVES TOO MUCH'

Stoljar's second argument against the revelation argument claims that the argument might prove too much. The thought here is that *if* there is an argument from revelation against grounding physicalism, then there is likely to be an argument from revelation

against Russellian monism. As Stoljar correctly observes, '[w]hile Russellian monism and physicalism are different, they have this in common (at least in their grounding versions): ordinary phenomenal properties are grounded in things that are not ordinary phenomenal properties'. This suggests, as Stoljar continues, 'that, if the revelation argument were successful against physicalism, it would be equally successful against Russellian monism'.¹⁰ While I agree with the latter assessment from Stoljar, I disagree that this shows that revelation is false or that the revelation argument 'proves too much'.

Let me first comment on the incompatibility between revelation and Russellian monism. Whereas grounding physicalism takes phenomenal properties to be grounded in physical properties, grounding versions of Russellian monism take phenomenal properties to be grounded in phenomenal or protophenomenal properties of states of fundamental micro-entities (Stoljar 2001, 2013b; Alter and Nagasawa 2012, 2015), or phenomenal or protophenomenal properties of states of the cosmos (Nagasawa and Wager 2016; Goff 2017, forthcoming).¹¹ In the previous section, we saw that if we formulate grounding physicalism by appealing to brute essentialism, then grounding physicalism is incompatible with revelation. With Russellian monism, the situation is the same. Someone who appeals to brute essentialism to formulate Russellian monism

¹⁰ Stoljar (2006, 2018) himself holds a panprotopsychoist version of Russellian monism. Stoljar thinks that this position avoids the problems of traditional physicalism and dualism. It does not face the conceivability argument, the knowledge argument or the combination problem faced by the panpsychist version of Russellian monism.

¹¹ Regarding grounding versions of Russellian monism, one might distinguish 'bottom-up versions' from 'top-down versions' (see Goff forthcoming). According to the former, it is facts on the bottom micro-level that ground facts on the higher levels. According to the latter, it is facts on the top macro-level that ground facts on the lower levels. Goff (2017, forthcoming) holds a top-down version of Russellian monism, known as 'cosmopsychism', according to which phenomenal facts are grounded in consciousness-involving facts about the cosmos.

as a grounding thesis would have at least to say that it is essential to phenomenal properties that they are grounded in some properties (which turn out to be (proto)phenomenal properties associated with fundamental entities or the cosmos), and these latter properties are such that if the relevant entities (either fundamental micro-level entities or the cosmos) are undergoing events with these properties, then subjects like us undergo experiences with phenomenal properties. However, this latter claim, which is similar to (G.ii*), is incompatible with the thesis of revelation. It just doesn't seem to be the case that my conscious experiences ever reveal that the phenomenal properties of my experiences are grounded in some other properties. So, for similar reasons, Russellian monism, formulated as a grounding thesis by appealing to brute essentialism, is also incompatible with revelation.

Our discussions in this and the previous section show that the thesis of revelation is in tension with many metaphysical theories of consciousness: revelation entails that identity physicalism is false; that role functionalism is false; and grounding physicalism and Russellian monism, given certain ways of developing the two, are also false. But the fact that the thesis of revelation is incompatible with these views does not, of course, on its own amount to an objection against the thesis of revelation itself. Were one of these theories to be non-negotiable or obviously true, one would have a reason to reject revelation. But it is far from clear that any of these theories is obviously true in the first place.

Note also that revelation does not rule out all the main contenders among metaphysical theories of consciousness. It certainly does not rule out property dualism,

according to which phenomenal properties are fundamental and the essences of phenomenal properties are not physical/functional. Note that revelation does not entail that, simply by having an experience with quale Q, one can know that 'the essence of Q is not physical/functional'. Revelation only entails that by having an experience with Q, one can know that 'Q is X', where 'X' captures the essence of Q. If the essence of Q were physical/functional, then 'X' would be a physical/functional predicate. Since, in having an experience, one is not in a position to know that Q is X where 'X' is a physical/functional predicate, one can then infer, from the thesis of revelation, i.e. from the knowledge that Q is X *and* that 'X' captures the essence of Q, that *the essence of Q* is not physical/functional. So, if one is committed to the thesis of revelation, one would also commit to dualists' claim that the essences of phenomenal properties are not physical/functional.

It could be that dualism, which is *compatible* with revelation, turns out to fare significantly worse than alternative metaphysical theories of consciousness which are *incompatible* with revelation, such as various versions of physicalism and versions of Russellian monism. So it may be that, despite the fact that revelation is often thought of as an intuitive claim, all things considered we have to deny revelation as a theoretical cost of preserving the best theory of consciousness. (I will turn to dualism in chapter 9.) But *prima facie*, whether or not we should reject revelation depends on whether there are good direct arguments against it. With that thought in mind, let's turn to Stoljar's objection to the thesis of revelation itself.

5. THE CASE OF MOORE

Stoljar's (2018) third objection against the revelation argument targets the thesis of revelation head on. Stoljar's objection is addressed specifically to Goff's formulation of revelation. Goff takes revelation to entail the claim that 'phenomenal concepts reveal the complete nature of the conscious states they refer to' (2017: 124). Stoljar's argument runs as follows. Take the phenomenal experience of pain. According to G. E. Moore, who held the sense-datum theory, to be in pain is to be acquainted with a particular kind of mind-dependent sense-datum.¹² Stoljar notes that most contemporary philosophers think that the sense-datum theory is false. Granted that it is false, Stoljar poses the following objection against revelation:

Now, that Moore held a false view of the complete nature of pain makes it plausible that he *failed to hold* the true view of the complete nature of pain, whatever that is. After all, given how rational Moore was (not to mention how interested in mental states he was), it is unlikely that he simultaneously held two inconsistent theories of pain. But then, according to the thesis of revelation, Moore did not have the phenomenal concept of pain. But that is absurd. Whatever the phenomenal concept is exactly, Moore is as good a candidate as any to possess it; it is just that he did not know the complete nature of pain.

We can understand Stoljar's objection as posing the problem that Goff's formulation of revelation entails that Moore has inconsistent beliefs. Moore's false belief in the sense-datum theory is supposed to show that he has a false belief about the nature of pain. But Goff's version of revelation, as Stoljar interprets it, plus the claim that Moore has the

¹² It is worth noting that the sense-datum theory is a theory about perception, in particular exteroception. Not all philosophers take pain experiences to be perceptual states. But sense-datum theorists usually think of pain experiences as paradigmatic cases in which the subject is related to the pain sense-datum (Broad 1925; Jackson 1976, 1977).

phenomenal concept of pain, entails that Moore has a true belief about the nature of pain. So, if Goff's version of revelation is true, then Moore has inconsistent beliefs about pain's nature. Since it is unlikely that Moore has two inconsistent beliefs, revelation – at least in the way that Goff formulates it – must be false.

One might think that Goff's formulation of revelation – 'phenomenal concepts reveal the *complete nature* of the *conscious states* they refer to' (2017: 124, italics added) – is either unclear or implausible. For instance, one might take it to be unclear whether Goff's version of revelation concerns conscious state-tokens or phenomenal properties of those states.¹³ One might take the term 'complete nature' to indicate that Goff is operating with a broad notion of essence that goes beyond the notion of immediate constitutive essence. Understood in that way, revelation does not seem to be particularly plausible. But even if we read Goff as advancing a version of revelation about the (immediate constitutive) essences of conscious state-types, i.e. phenomenal types or phenomenal properties, one might still think that his formulation is unnecessarily strong as a claim about concept possession. It states that whenever one possesses a phenomenal concept that refers to phenomenal property Q, one automatically knows the essence of Q.

¹³ Elsewhere, it does seem that Goff (2017: 107; 2015: 214) intends his formulation of revelation to be about conscious state-types, i.e. phenomenal types.

In the following, I will focus on the formulation of revelation proposed in chapter 4. I will assess whether Stoljar's objection, once adjusted, problematises that formulation.¹⁴ Recall that the thesis of revelation says:

(Revelation)

By having an experience-token with phenomenal property Q, one is in a position to know a truth, namely, 'Q is X', where the predicate 'X' captures the essence of Q, although it may be hard to put into words.

Let Q be the phenomenal property of a pain experience, i.e. the painfulness of pain. Given revelation, if Moore had a true belief that was explicitly about the nature or essence of quale pain, i.e. a belief that the *essence* of quale pain is thus-and-so, then Moore would believe, not just 'Q is X', but the following:

(a) Q is X and 'X' captures the essence of Q.

We also know that Moore holds the sense-datum theory. According to the sense-datum theory, in veridically perceiving something red and in hallucinating something red, the subject of the experience is related to a red-sense-datum.¹⁵ On this picture, one might say

¹⁴ It has been suggested to me in conversation with Stoljar that he does not take his argument against revelation to depend on Goff's specific formulation. So, one can reasonably assume that the case of Moore targets the version of revelation formulated in chapter 4.

¹⁵ Sense-datum theorists adopt the questionable *phenomenal principle*, according to which in having an experience, whether veridical or hallucinatory, if one is aware of something having a sensory property F, then there must be something that one is aware of which has F (Robinson 1994). But in cases of illusion and hallucination, while one is aware of something having F, there is no mind-independent object that has F. This leads sense-datum theorists to posit mind-dependent entities, i.e. sense-data, which can serve as bearers of the relevant property. Sense-datum theorists also hold the 'common kind assumption', according to which veridical perceptual experiences on the one hand, and corresponding hallucinations or illusions on the other hand, are of the same kind or have the same nature (Martin 2002).

that the phenomenal property of phenomenal red is determined by the property of the particular sense-datum given in experience, which presumably has a certain red-like quality. One might then spell out the essences of phenomenal properties in the following manner and take Moore, who is a sense-datum theorist, to believe (b):

- (b) Q is *essentially* such that if one's experience is a relation to a particular sense-datum Q*, then one's experience has phenomenal property Q.

Let's grant for now that a sense-datum theorist specifies the essence of Q, i.e. the essence of the phenomenal property of the pain experience, in this manner. Following Stoljar, let's also grant that the sense-datum theory is false. So, Moore *falsely* believes the sense-datum theory to be true – his belief in (b) is false.

Note that belief in (a) is straightforwardly inconsistent with belief in (b). If the thesis of revelation is true, then the predicate 'X', in the essence-capturing truth 'Q is X', which the subject of an experience with Q is in a position to know, cannot be a predicate that mentions a 'sense-datum'. Experience itself does not reveal to us anything about sense-data. So one cannot consistently believe (a) and (b) simultaneously – it would amount to believing two incompatible views of the essences of qualia.

In Stoljar's original objection, Goff's version of revelation is problematic because it has the undesirable consequence that Moore has inconsistent beliefs about the nature of pain, i.e. the essence of the pain phenomenal type (quale pain). Given our formulation of the thesis of revelation, we can now reformulate Stoljar's objection as the following:

- (1) If the thesis of revelation is true, then Moore has a true belief about the essence of quale pain, i.e. (a).¹⁶
- (2) Moore has a false belief about the essence of quale pain, i.e. (b).
- (3) If revelation is true, then Moore has inconsistent beliefs.
- (4) Moore does not have inconsistent beliefs.
- (5) Revelation is false.

I shall argue that there are various moves available to a proponent of revelation in responding to the above argument, and that even if a proponent of revelation makes various concessions to her opponent, Stoljar's case of Moore still fails to raise a forceful objection against revelation.

Denying (1): The objection does not go through because (1) is false. The thesis of revelation, given the current formulation, merely claims that experience puts one *in a position* to believe or know (c):

- (c) Q is X (where 'X' captures the essence of Q).

Two points are worth clarifying. First, while Moore is *in a position to know* (c), he might fail to believe (c), and instead believe (b), that Q's essence has something to do with some

¹⁶ A cautious reader might already notice that the thesis of revelation does not entail (a). It says only that the subject is in a position to know that Q is X (where 'X' does, in fact, capture the essence of Q). My point here is that in order to get Stoljar's objection started, we need to take Moore to have a true belief about *the essence of Q*, i.e. (a), and also to have a false belief about *the essence of Q* being something about sense-data, i.e. (b). The issue here will be clarified below.

sense-datum. Second, even if Moore believes (c), (1) is still false because (c) is different from (a) and is not in conflict with (b). Recall (a):

(a) Q is X *and* 'X' captures the essence of Q.

It is possible that one believes (c) without believing (a). Moore might believe that 'Q is X' while at the same time thinking that the predicate 'X' does not capture, i.e. describe accurately, the essence of Q.

Whereas belief in (b) is inconsistent with belief in (a) as we saw, it is not inconsistent with belief in (c). A subject S can consistently believe that 'Q is X' (where 'X' does in fact capture the essence of Q), and at the same time believe that 'Q is essentially Y', where the predicate 'Y', which is distinct from 'X', is believed by S to capture the essence of Q. So it could be the case that Moore believes (c), not (a), but at the same time also believes (b), and there is nothing inconsistent about believing (b) and (c) simultaneously.

Furthermore, it seems plausible to say that even if we grant that Moore already believes (c) and that he has reflected hard on the nature of phenomenal properties, it may still be the case that he has failed to believe (a). One might think that many physicalists are in the same boat as Moore. They have thought hard about the essences of qualia; they can even agree that by having an experience with phenomenal property Q, one can arrive at the belief that 'Q is X' where 'X' specifies the what-it-is-likeness of Q. But such physicalists, *contra* advocates of revelation, insist that 'X' does not capture, that is, characterise accurately, but merely *refers* to the essence of Q, and that Q has a hidden essence, namely, being a certain physical property (Loar 1997; Papineau 2007;

Balog 2012). So, it could very well be the case that Moore, like these physicalists, tends to think that the essence or nature of the phenomenal character of experience is *not* revealed in experience. Whereas further theorising leads physicalists to hold the view that qualia have physical or functional essences, it leads Moore to believe in the sense-datum theory. So, the argument fails to go through because Premise (1) is false.

Denying (2): For the sake of argument, let's grant our opponent the (incorrect) claim that revelation entails that Moore believes (a), thus granting premise (1). But premise (2) is doubtful. It is questionable that one needs to commit to (b) in order to uphold the sense-datum theory. The sense-datum theory is first and foremost a theory about the structure of perceptual experiences; it is not put forward as a theory about the essences – still less, as a theory specifically about the immediate constitutive essences – of phenomenal properties of perceptual experiences. In contrast, revelation is not a claim about the structure of experiences (i.e. experiential events); but a claim about the essence (that is, the immediate constitutive essence) of a phenomenal property of a given experience. The sense-datum theory says that a token perceptual experience at a given time is an instantiation of an acquaintance relation between a subject and a sense-datum. Accordingly, a particular structural type of perceptual experience, E, can be conceived as essentially a relation of acquaintance between the subject and a particular type of sense-datum (e.g. Q*). But token experiences of type E might also have qualia like Q, and in addition, Q is such that by having an experience with Q, one is in a position to know the truth 'Q is X', where 'X' captures the essence of Q. It might be due to a nomic law or a metaphysical law that phenomenal properties like Q are correlated with sense-data

like Q^* . On this picture, being correlated with the sense-datum Q^* is, intuitively, merely a nomically or metaphysically necessary feature of Q , not part of the latter's essence in its most core respects. In other words, the sense-datum Q^* should not feature in the real definition of Q in the fashion of (b). So, premise (2) can be rejected and the argument does not go through.

However, one might think that although Stoljar appeals to the case of Moore to problematise revelation, the example of Moore is merely a stand-in. One can equally construe a case where the protagonist is presumably rational but holds a false view of the essences of qualia that is incompatible with the thesis of revelation. So, let's grant that being a sense-datum theorist, Moore is committed to (b). In other words, we are granting our opponent both the incorrect premise (1) and the (otherwise unmotivated) premise (2); and thus granting step (3) of the objection.

Denying (4). Even if we grant (1) and (2) in order to explore the objection more fully, a proponent of revelation still has a response to the above argument against the thesis of revelation. One can deny (4) and concede that Moore has inconsistent beliefs about the essence of quale pain.

Here one can maintain that this kind of inconsistency is not problematic. A thinker as rational as David Lewis (1982: 436) admits to having had beliefs that were inconsistent – his system of beliefs was 'fragmented' and different fragments 'came into action in different situations'. Indeed, this phenomenon is entirely normal and such fragmentation is crucial to allowing beliefs to be evaluated after their initial adoption

even without new evidence (see Egan 2008; Davies and Egan 2013). Given these points, Moore's inconsistency cannot be used as the key step in an argument against revelation.

Furthermore, certain ordinary beliefs may be especially tenacious and resistant to revision. Consider Cicero (2006: 38), who called himself an 'Academic Sceptic', someone who holds the belief that there is no such a thing as knowledge. But Cicero recognises that he does not act in accordance with this belief, and in everyday contexts, he acts as if he believes that he has knowledge. It seems that it may be very difficult to overcome one's predisposition to maintain certain ordinary beliefs and to revise them in order to avoid inconsistency with one's theoretical stance. The thesis of revelation may be such a belief, as many physicalists indeed think (1995; McLaughlin 2003; Braddon-Mitchell 2007; Hill 2014), and the idea that experience affords us knowledge about the essences of qualia, as we saw in chapter 6, is plausibly a deep part of our ordinary thinking.

It might be that Moore struggles to overcome the inconsistency in his belief system; or it might be that he alternates between believing (a) and believing (b), never actually holding inconsistent beliefs simultaneously. But it could well be that Moore's case is similar to Lewis' (1982: 436) fragmented system of beliefs, where 'different fragments came into action in different situations, and the whole system of beliefs never manifested itself all at once'. Perhaps when Moore takes off his sense-datum theorist's hat, what he says and does is guided by his belief (a), that Q is X and 'X' captures the essence of Q. But when he is in a metaphysical-theoretical mindset, he is guided by his belief (b) instead. Whatever the details of the case, the upshot is that we need not jettison

the thesis of revelation just because Moore holds inconsistent beliefs about the essences of qualia.

Let's now summarise our discussion in this section. Recall that Stoljar's objection against the thesis of revelation, based on the case of Moore, targets Goff's formulation of the thesis specifically. But an advocate of revelation need not adopt Goff's formulation of the thesis. In assessing whether Stoljar's case of Moore poses a problem for revelation, we have adapted Stoljar's argument to bring it into contact with the formulation of revelation laid out in chapter 4. As we saw, there are various moves a proponent of revelation can make in response to the adapted case of Moore. We saw that even when a proponent of revelation makes substantial concessions to her opponent (granting premises that are incorrect or unmotivated), the thesis of revelation leads at most to the claim that someone as rational and thoughtful as Moore might have held inconsistent beliefs. So, far from being a *reductio ad absurdum* of the thesis of revelation, this places Moore in the company of Cicero, Lewis, many philosophers who have thought deeply about physicalism, and, indeed, every normal person who compartmentalises a hastily adopted belief so that it can be evaluated when time and other resources allow. Overall, it is far from clear that Moore's case, as we have adapted it from Stoljar's (2018) original, poses a problem for the thesis of revelation.

Indeed, it would be surprising if the thesis of revelation were defeated by Moore's false belief in the sense-datum theory. Revelation by no means implies that there could be no false metaphysical theories about the structure of token experiences. Nor does it imply that there can be no false theories or disagreements about the essences of

qualia. In introspecting the phenomenal character of an experience, e.g. phenomenal red, advocates and deniers of revelation can both come to the phenomenal knowledge that 'phenomenal red is (like) *this*', where the demonstrative 'this' is a placeholder for the what-it-is-likeness of phenomenal red. An advocate of revelation takes the demonstrative 'this' to *reveal* the essence of the property phenomenal red and thus takes herself to know what makes phenomenal red the property it is. In contrast, a denier of revelation merely takes the demonstrative 'this' to *refer* to the essence of phenomenal red (e.g. some physical property of being a neural event of a particular type, or some relation to a sense-datum of a particular type) which she takes herself to be ignorant about in having an experience with the property phenomenal red. Such a disagreement about whether the thesis of revelation is true is not an introspective disagreement, but arises from different theoretical commitments. Someone who is committed to a reductionist physicalist worldview is unlikely to think that what we know about qualia through introspection reveals the essences of qualia. But, of course, the fact that one holds a theory about qualia that is incompatible with revelation is not, by itself, adequate reason to reject the thesis of revelation.

One might, however, question whether it is the job of introspection to reveal the essences of qualia. In the following section, I shall turn to a potential worry regarding revelation that is related to the limits of introspection.

6. INTROSPECTIVE HUMILITY

In section 1 of this chapter and also in section 4 of chapter 4, I mentioned that in the literature the thesis of revelation is often taken to be a claim about introspection, that *introspection reveals the essences of qualia*. Given our formulation of revelation, we can formulate the latter claim in the following way:

- (I) Introspection reveals that Q is X: By introspecting an experience-token with phenomenal property Q, one is in a position to know that Q is X, where 'X' captures the (immediate constitutive) essence of Q.

For the purposes of this section, (I) will be taken to be equivalent to the thesis of revelation.

In their paper 'Introspective Humility', Tim Bayne and Maja Spener (2010) argue that while introspection is largely reliable, it goes awry when it is employed outside its operational constraints. Operational constraints are constraints on the proper functioning of introspection, including that the underlying mechanisms are intact, introspection is being deployed under appropriate conditions, etc. (2010: 10). Bayne and Spener contend that the fact that introspection is being employed outside its operational constraints partially explains the presence of introspective disagreements with respect to certain issues which pertain to domains that fall outside introspection's proficiency.¹⁷

¹⁷ According to Bayne and Spener (2010: 9), introspective disagreements can also be due to individual differences, terminological variations and differences in background beliefs and expectations.

Two cases of introspective disagreement Bayne and Spener discuss that are of interest here are the debate about *cognitive phenomenology* and that about *the nature of the phenomenal character of perceptual experience*. Bayne and Spener contend that introspection is not reliable at revealing whether thoughts have *sui generis* phenomenology, or whether the phenomenal character of perceptual experience has aspects that are non-representational. In the following, I will look into these two debates and clarify whether the limits of introspection with respect to these issues cast any doubt on the thesis of revelation. I shall argue that they do not.

Let's consider the debate about cognitive phenomenology. This debate is fought between those who think that thoughts have a *sui generis* phenomenology and those who think they do not. Bayne and Spener (2010: 4-5) note that philosophers involved in this dispute disagree at the level of introspection – while some take introspection to reveal that cognitive phenomenology is sensory, others take it to reveal that it is *sui generis* (see Bayne and Montague 2011; Spener 2011).¹⁸

Take a particular cognitive state with phenomenal character C whose essence is Y. Given (I), we have the following claim:

¹⁸ They cite Horgan and Tienson (2002) as appealing to introspection to support the idea that cognitive phenomenology is *sui generis*. They cite Wilson (2003) as arguing that introspection reveals the opposite. One might question whether there is actually a disagreement over what introspection alone reveals. Elijah Chudnoff (2015) notes that theorists who are unbiased, terminologically calibrated, etc. can indeed agree on what introspection reveals, e.g. 'I can immediately tell whether I am consciously thinking that 3 is prime.', 'There is a felt difference between hearing "Dogs dogs dog dog dogs" as a mere list and hearing it as a meaningful sentence.', etc. But due to differences in argumentation and theorising, they take such introspective evidence to show different theories about cognitive phenomenology.

(I') Introspection reveals that C is Y: By introspecting a cognitive state with phenomenal property C, one is in a position to know that C is Y, where 'Y' captures the essence of C.

Granted the presence of introspective disagreement with respect to cognitive phenomenology, it seems:

(CP) It is not true that introspection reveals whether C is sensory or *sui generis*.

Is there any tension between (CP) and (I')? Put differently, is the following claim plausible?

(ϕ) If introspection reveals that C is Y, where 'Y' captures the essence of C, then introspection reveals whether C is sensory or *sui generis*.

I shall argue that there is no tension between (I') and (CP) because an advocate of revelation need not hold (ϕ).

We can draw a distinction between, on the one hand, knowing that C is Y, where 'Y' captures the essence of C, and, on the other hand, knowing what category *quale* C belongs to or what natural kind C is an instance of. Knowing that C is Y (where 'Y' captures the essence of C) can be achieved by knowing that C is 'this', where the demonstrative, which refers to the what-it-is-likeness of C, may be difficult or impossible to unpack in further words. With this knowledge, one might come to possess a concept of the phenomenal property C, be able to remember that property, imagine that property,

recognise further instances of that property, and so on. Knowing what category quale C belongs to is more demanding than knowing that quale C is Y (where 'Y' captures the essence of C). It concerns categorising and locating a particular quale using a particular taxonomy. It requires one to have some level of mastery of the taxonomic system at issue, to know about different categories of qualia, how they differ from one another, etc. Consider the analogy of knowing colours. On a primitivist conception of colour properties, where colours are intrinsic, qualitative and irreducible properties of objects, one might say that knowing the essence of the colour scarlet does not entail that one also knows taxonomic truths about scarlet, e.g. scarlet is a shade of red, not a shade of yellow, etc. (see Campbell 2005). Imagine someone who since birth has been trapped in a world where everything is scarlet. On the colour primitivist account, this person knows exactly what scarlet is without knowing any taxonomic truths about it. Equally, one might know all taxonomic truths about scarlet, e.g. knowing scarlet is a shade of red, not a shade of yellow, etc., without knowing what scarlet really is or what it looks like.

Something similar might be going on in the case of cognitive phenomenology. I might know exactly what the phenomenology of having a certain thought is when I have that thought. But this does not entail that I know that the phenomenology of thinking is sensory or know that it is *sui generis*. I might simply not have sufficient knowledge about the taxonomic system in which I need to locate the phenomenology of thinking. The introspective disagreement about cognitive phenomenology that Bayne and Spener discuss in their paper concerns the latter kind of knowledge. Such an introspective disagreement, granted that there is one, could mean that introspection is not able to

reveal certain taxonomic truths about phenomenal properties. But knowing those taxonomic truths does not amount to knowing the essence of a phenomenal property (e.g. knowing that C is Y, where 'Y' captures the essence of C). Nor does knowing the essence of a phenomenal property entail that one also knows taxonomic truths about that phenomenal property. Therefore, since we need not hold (ϕ), there is no tension between introspective disagreements about cognitive phenomenology or (CP) on the one hand, and (I) or the thesis of revelation on the other.

Bayne and Spener (2010) also discuss the debate regarding the metaphysical structure of the phenomenal character of perceptual experience, fought between 'those who hold that the phenomenal character of perceptual experience is exclusively world-directed' and 'those who hold that such character involves aspects that are not world-directed' (2010: 3). They (2010: 3) note that 'in contemporary discussion this division fuels the debate between pure representationalists ... and impure representationalists'. Both pure and impure representationalists about perceptual experience agree that all perceptual experiences are representational. But while pure representationalists hold that the phenomenal character of a perceptual experience is exhausted by (or wholly determined by) the representational content of the experience, impure representationalists think that the phenomenal character of a perceptual experience also

has aspects that are non-representational (see Crane 2000).¹⁹ Bayne and Spener contend that this disagreement also occurs at the level of introspection.²⁰

Take a perceptual experience with phenomenal character P, where P's essence is Z. Given (I), we have the following claim:

(I'') Introspection reveals that P is Z: By introspecting a perceptual experience with phenomenal property P, one is in a position to know that P is Z, where 'Z' captures the essence of P.

Granted that there is a genuine introspective disagreement regarding the metaphysical structure of the phenomenal character of perceptual experience, it seems (PE) is true:

(PE) Introspection does not reveal whether or not P has aspects that are non-representational.²¹

Now, is (I'') in any tension with (PE)? Put differently, is the following claim plausible?

¹⁹ By 'representational properties', I here mean what Chalmers (2010a) calls 'pure representational properties', which are properties of representing certain intentional contents. Such properties are contrasted with so-called 'impure representational properties', i.e. properties of representing certain intentional contents in certain ways (see chapter 2 footnote 9).

²⁰ As examples, Bayne and Spener (2010: 3) cite Levine (1995: 277-8) as appealing to introspective evidence to support his position that (perceptual) qualia are non-representational in the sense that they cannot be given a relational analysis, e.g. functional/representational analysis. They (2010: 3) also cite Tye (1992: 60) as resorting to introspection to support pure representationalism about perceptual qualia. For Levine, but not for Tye, perceptual experience also has aspects that are non-representational. Whether Bayne and Spener's exegeses of Levine and Tye are correct is an issue I won't delve into here, though I think there are good reasons to question their exegeses.

²¹ One might question whether there is a genuine disagreement at the level of introspection with respect to the metaphysical structure of the phenomenal character of perceptual experience. It may well be that the disagreement is not at the level of introspection, but results from differences in theorising. The argument in the main text is intended to show that, *even if* the disagreement *is* at the level of introspection, it does not challenge the thesis of revelation.

(ψ) If introspection reveals that P is Z, where 'Z' captures the essence of P, then introspection reveals whether or not P has aspects that are non-representational.

I shall argue that there is no tension between (PE) and (I'') because an advocate of revelation need not hold (ψ).

Let's first compare (ψ) with the following claim:

(ω) If introspection reveals that Q is X, where 'X' captures the essence of Q, then introspection reveals whether or not Q is physical/functional.

In chapter 4, we saw that the thesis of revelation is incompatible with standard versions of physicalism according to which phenomenal properties are physical/functional. Thesis (I) (near the beginning of this section) is also incompatible with standard physicalism. The incompatibility can be spelt out as follows: Given (I), if standard physicalism is true, then introspection would reveal that Q is X, where 'X' is some physical/functional predicate; but introspection does not reveal Q to be X, where 'X' is some physical/functional predicate; thus, standard physicalism is incompatible with (I). So, on the basis of (I), one can conclude, through *inference*, that qualia are not physical/functional.

Importantly, this is not to say that *introspection* reveals that qualia are not physical/functional. Introspection, by itself, only reveals the essences of qualia, e.g. Q is X where 'X' captures the essence of Q, and the conclusion that qualia are not physical/functional is arrived at through further reasoning. Thus, (ω) is *to be rejected* – it

is not the case that if introspection reveals the (immediate constitutive) essences of qualia, then introspection also reveals whether or not qualia are physical/functional.

Similarly, we can also reject (ψ) and say that it is not the case that if introspection reveals the essences of perceptual qualia, then it reveals whether or not perceptual qualia have aspects that are non-representational.²² The picture I am proposing is this. Suppose that one has a perceptual experience of, or as of, some worldly scene. This token experience has various phenomenal properties and either (i) all these phenomenal properties are exhausted (or wholly determined) by the representational properties of the experience, i.e. properties of representing the world to be thus-and-so, or else (ii) some of these phenomenal properties have aspects that are non-representational, in which case the phenomenal character of the experience is not exhausted by its representational content.²³ Consider some quale Q , which is a complex phenomenal property. One can know that Q is X (perhaps using a kind of demonstrative), where ' X ' captures the essence of Q , without knowing whether Q has aspects that are non-representational. Similarly, if Q is a simple phenomenal property, one can know that Q is X , and possess a phenomenal concept of Q , without being able to theorise whether Q

²² One can nevertheless maintain that what one knows in knowing the essence of Q by having an experience with Q , i.e. Q is X (where ' X ' captures the essence of Q), provides the resources to know whether Q has aspects that are non-representational. Of course, it is difficult to reach a clear verdict on the question of whether perceptual qualia can ever have aspects that are non-representational. To know the latter requires one to at least apply the concept of being representational and should introspectively examine a vast number of token experiences.

²³ One candidate example is the experience of blurry vision. In the case of blurry vision, one experiences blurrily so to speak. Pure representationalists would say that experience in this case represents the world as blurry. However, one might draw a phenomenological difference between seeing blurry things, i.e. things with fuzzy boundaries, and seeing things in a blurry way (see Crane 2000).

is representational or non-representational. If this is right, then (ψ) , like (ω) , is *to be rejected*. It is not true that if introspection reveals the essences of qualia, then introspection reveals whether qualia have aspects that are non-representational.

Here it may also be helpful to return to Fine's distinction between immediate essence and mediate essence, which was introduced in section 2 of chapter 4. Consider this example. Intuitively, in knowing the truth 'Water is molecules made up of two hydrogen atoms and one oxygen atom', one knows the essence, that is the immediate (also constitutive) essence, of water, because the statement captures the essence of water in its core respects (see Fine 1995a, 1995b; Dasgupta 2014: 589). To know the essence of water in this sense does not require one to know what the correct theory of the atom is or what the essence of an atom is. To know the latter would be to know what belongs to the mediate essence of water.

Similarly, in knowing that the phenomenal property Q is X, one knows what is in fact the essence, i.e. the immediate constitutive essence, of Q – even if one does not know *that X is the essence of Q*. If one also knows whether Q is wholly representational or has aspects that are non-representational, what one knows might be part of the mediate essence of Q. Since we are only concerned with the immediate constitutive essences of phenomenal properties, it is safe to say that one can know the essence, i.e. the immediate constitutive essence, of Q, without knowing whether or not Q has aspects that are non-representational. So (ψ) is false.

Our discussion on introspective humility shows that the limits of introspection which, according to Bayne and Spener (2010), may give rise to introspective

disagreements in areas like cognitive phenomenology and the metaphysical structure of the phenomenal character of perceptual experience, pose no threat to the thesis of revelation. An advocate of revelation can maintain that introspection reveals the essences of qualia without revealing whether qualia are sensory or *sui generis*, or whether they have aspects that are non-representational.

7. CHAPTER CONCLUSION

In this chapter, I surveyed potential worries for the thesis of revelation and considered in depth some objections against the revelation argument against physicalism. Stoljar's third objection against the revelation argument is particularly notable, as it targets the thesis of revelation directly. I showed that the adapted objection from Stoljar against my formulation of the thesis of revelation fails, and argued that Stoljar's two other objections regarding the revelation argument against physicalism are not convincing either. In addition, I related the thesis of revelation to the discussion on introspection. I argued that even if introspection cannot reveal whether a certain quale is sensory, representational, etc., this does not mean that introspection cannot reveal the essence of that quale.

As we saw in this chapter, the thesis of revelation does not entail that by having an experience with phenomenal property *Q*, one is in a position to know the metaphysical structure of the experience, or all the philosophically important features of *Q*, e.g. whether it is sensory, or representational, etc. Not knowing these latter claims does not prevent one from knowing *Q*'s immediate constitutive essence. But the latter

knowledge does put us in a position to see that standard physicalism, which identifies phenomenal properties like Q with physical/functional properties, is false outright, and that grounding versions of physicalism and Russellian monism might also be false.

Given the discussion in this chapter and chapters 4, 5 and 6, we can conclude that the thesis of revelation, although incompatible with many metaphysical theories of consciousness, is nevertheless an intuitive claim and is not vulnerable to serious direct objections.

Chapter 9

REVELATION AND DUALISM

1. WHAT WE HAVE ACHIEVED SO FAR

Throughout the dissertation, we have seen various ways to draw distinctions between different kinds of physicalism. We saw (chapter 2) the distinction between identity physicalism and functionalism, the distinction between reductive physicalism and nonreductive physicalism, and that between a priori physicalism and a posteriori physicalism. While there may be various ways to draw distinctions within physicalism, physicalists standardly claim that the phenomenal properties of our experiences are just physical or functional properties. But it seems that no matter how much we know about the physical processes underlying our conscious experiences, we are still left wondering how the phenomenal character of our experience – what experiences are like for the subject – could be physical/functional. This persistent puzzlement has led physicalists like Papineau (2002, 2007, 2011, forthcoming) to say that our resistance to physicalism stems from a deeply ingrained intuition of dualism, an intuition that consciousness is nonphysical, which can be explained in a way that is consistent with the truth of physicalism (see also Chalmers 2018).

This dissertation has addressed the intuition of dualism and its explanation directly. I think that the existence of a dualistic dualism is undeniable and that a deeper understanding of it can guide our philosophical theorising about the metaphysics of

consciousness. Here is a summary of what we have achieved so far in the course of the previous eight chapters. In providing this summary, I shall also note several points that call for further research.

In Part I of the dissertation, I motivated the claim that there is an intuition of dualism and surveyed several candidate explanations for the intuition. In chapter 2, I argued that the intuition of dualism is made evident by a closer look at the problem of the explanatory gap levelled against physicalism. As we saw, physicalists' strategies for bridging the explanatory gap seem unsatisfactory because phenomenal-physical identity claims such as 'the experience of pain (in humans) is C-fibres firing in the brain' are disanalogous to familiar scientific identity claims like 'water is H₂O'. Notably, while we do not persist in wondering why water is H₂O or why H₂O is water, we continue to feel puzzled about how our experience of pain could just be C-fibres firing and how C-fibres firing could amount to the experience of pain. I argued that this disanalogy makes it evident that there is a persistent, and perhaps a widespread, intuition of dualism, an intuition that conscious experience or consciousness is nonphysical.

Chapter 3 laid out the desiderata for an adequate explanation of the intuition of dualism. I clarified that such an explanation can take various forms – it can be rational, mechanistic, historical/cultural, and so on – but it must address the target intuition and predict correctly the range of cases where the intuition of dualism arises. I then critically surveyed three notable proposals: Fiala, Arico and Nichols' (2011) proposal which appeals to the dual-process cognition that allegedly underlies our third-person attributions of conscious states; Nagel's (1986) idea that the subjective and the objective

do not seem to converge; and Papineau's (2002) account which appeals to the uniqueness of phenomenal concepts. I argued that these proposals either fail to explain the target intuition or fail to predict correctly cases where the intuition of dualism arises. The question of what explains the intuition of dualism was thus left open at the end of Part I.

In Part II of the dissertation, I put forward a rational explanation for the intuition of dualism, appealing to the intuitiveness of the thesis of revelation, or, more precisely, to our tacit understanding of the thesis of revelation, which forms part of our ordinary conception of experience. Much of Part II was aimed at clarifying the thesis of revelation in the philosophy of mind. Conscious experiences have specific phenomenal properties which have essences, i.e. *immediate constitutive essences* in the sense of Fine (1995a, 1995b). The essence of a phenomenal property distinguishes it from other phenomenal properties. According to the thesis of revelation, as we saw in chapter 4, by having an experience with a specific phenomenal property, one is in a position to know the essence of that property. I spelt out this knowledge as a kind of knowledge *de dicto*, e.g. 'Q is X', where 'X' is a predicate that describes accurately what Q is in its most core respects, though such a predicate may be hard to put into words. The thesis of revelation, as I clarified, is a knowledge-attribution *de dicto* of a kind of knowledge *de dicto*.

There are a number of conclusions one can draw from the thesis of revelation thus understood. One can plausibly draw the conclusion that some qualia are simple. Regarding the phenomenal properties of some experiences, it is not the case that we notice any internal structures of these properties. On the assumption that internal

structures would belong to the essences of phenomenal properties, it follows that these phenomenal properties are simple.

From the thesis of revelation, one can also conclude that it is false that phenomenal properties have physical/functional essences, i.e. standard versions of physicalism are false. This was addressed in chapter 5. Identity physicalism claims that phenomenal properties are physical properties, e.g. the painfulness of a pain experience is the property of being an event of C-fibres firing. It is obvious that revelation is incompatible with identity physicalism. By having a pain experience, we are certainly not in a position to know anything about C-fibres firing. I also argued that revelation is incompatible with role functionalism, in particular commonsense role functionalism. I argued that commonsense role functionalists who claim that their position is compatible with revelation face a dilemma. If they define qualia in functional terms by appealing to everyday concepts of worldly objects and states like 'British pillar box' in the case of phenomenal red (see Lewis 1997: 335), then their position is incompatible with revelation. But if they avoid making reference to these things in spelling out the commonsense functional roles of qualia, then it seems no longer clear that, with only the vocabulary that would not create a problem for maintaining revelation, they can give adequate functional definitions that individuate qualia. Given this dilemma, we also have an argument from revelation against commonsense role functionalism. The thesis of revelation also seems to be incompatible with grounding physicalism and grounding versions of Russellian monism (see chapter 8).

Though the thesis of revelation is in tension with various metaphysical theories of consciousness, it nevertheless is an intuitive claim. This was argued for in chapter 6 on the basis of certain linguistic considerations, focusing in particular on the oddness in sentences such as ‘I know what an itch feels like, but I don’t know what the feeling of an itch really is’. I argued that the relevant linguistic data are best explained by the hypothesis that the thesis of revelation is part of our ordinary conception of experience. By inference to the best explanation, we can conclude that the thesis of revelation is, indeed, part of our ordinary conception of experience, and can be said to be intuitive in this sense.

The discussions in chapters 4, 5 and 6 on the thesis of revelation allowed us to put forward a rational explanation for the intuition of dualism in terms of the intuitiveness of revelation in chapter 7. According to this explanation, our intuition of dualism has an inferential aetiology that can be traced to the thesis of revelation. Our judgement or disposition to judge that consciousness is nonphysical draws on our tacit understanding of the thesis of revelation and a tacit appreciation of the entailment from revelation to the claim that phenomenal properties are not physical properties (including functional properties). By the end of Part II, the question left open at the end of Part I, namely, the question of what explains the intuition of dualism, had been answered and the main dialectical arc of the dissertation completed.

In chapter 7, I explored two additional issues, each of which could lead to a substantial programme of further research. First, I considered the possibility of a similar intuition of distinctness with respect to colour properties. The conclusion I drew there

was conditional: *if* the thesis of revelation about colour properties, i.e. revelation-c, is deeply rooted in our ordinary thinking, then there is a similar intuition of distinctness with respect to colour properties. While some philosophers think that only phenomenal properties, not colour properties, generate an intuition of dualism (e.g. Chalmers 2018), I think the issue here is less obvious (see Byrne 2006; Sundström 2007). An extended comparison of the two theses of revelation would be a worthwhile project.

The issue here also turns on wider debates about theories of perception. Naïve realism and revelation-c are often thought of as natural bedfellows (see Allen 2011, 2017) – those who hold revelation-c are usually naïve realists about perception. In contrast, disjunctivism about phenomenal properties, which is commonly endorsed by naïve realists, is incompatible with the thesis of revelation about phenomenal properties (revelation-q) – it is not the case that in having an experience with phenomenal property Q, we are in a position to know that Q is X where ‘X’ is a disjunctive predicate that captures the essence of Q (which is disjunctive). I think a further in-depth comparison of the two theses of revelation would shed light on not only the problems faced by the physicalist ontology, but also the debate about the metaphysical structure of perceptual experiences.

The second additional issue that I discussed in chapter 7 was the question of why revelation is part of our ordinary conception of experience. There, I discussed two candidate explanations. I outlined how the truth of revelation itself would provide an explanation for the intuitiveness of revelation. Then, I critically assessed a putative physicalist explanation for the intuitiveness of revelation that appeals to Pereboom’s

(2011) qualitative inaccuracy hypothesis. I argued that this hypothesis, given its two possible readings, is either implausible or fails to provide a satisfactory explanation for the intuitiveness of revelation.

My discussion in chapter 7 leaves it open that there could be, besides Pereboom's, other physicalist proposals for explaining away the intuitiveness of revelation. The issue here is related to the recent surge in discussion on consciousness from the illusionists' camp to which Pereboom belongs (Frankish 2016; Chalmers 2018; Kammerer 2018b, forthcoming). Recall (from chapter 7, sections 5 and 6) the two components of the thesis of revelation:

- (i) By having an experience with phenomenal property Q,
one is in a position to know *de dicto* that Q is X.
- (ii) 'X' captures the essence of Q.

Understood in a certain way, (i) seems very compelling regardless of whether or not the thesis of revelation is true: by having an experience with quale Q, one is in a position to know that Q is X, where 'X' captures the *what-it-is-likeness* of Q (though this might be hard to put into words). My position is that given (i) thus understood, it is easier to motivate the claim that the essence of Q is revealed in experience, hence (ii), than to argue that it is hidden. But some physicalists would of course disagree and deny (ii) in spite of agreeing to (i) where 'X' captures the what-it-is-likeness of Q. Illusionists like Pereboom take a more radical approach. They maintain that (i) (where 'X' captures the what-it-is-likeness of Q) is in fact erroneous, and consequently reject (ii) and the thesis of revelation. But the denial of (i) (where 'X' captures the what-it-is-likeness of Q) is

highly counterintuitive and calls for an explanation. Illusionists need to explain why it is the case that (i) seems compelling when it is in fact false. They need to explain this in a way that does not make their proposal, like Pereboom's, unclear or unmotivated or downright implausible. On the face of it, I am not sure how such an explanation would go. Nonetheless, it is worth engaging further with illusionists in future research (e.g. Frankish 2016; Kammerer forthcoming).

Part III goes further beyond the dissertation's main narrative of explaining the intuition of dualism. Chapter 8 turns to the force of the revelation argument against physicalism and in particular the question of whether the thesis of revelation is in fact true. I addressed two existing objections against revelation – Trogon's claim (2017) that the thesis of revelation is not a plausible view, and Stoljar's objection (2018) that revelation entails that someone as rational and thoughtful as Moore holds inconsistent beliefs. In addition to defending the thesis of revelation from these existing objections, I also turned to discussion on the limits of introspection. I argued that the relevant limits on the proper functioning of introspection – namely, that introspection does not reveal whether cognitive phenomenology is *sui generis* or sensory or whether the phenomenal character of perceptual experience has aspects that are non-representational – does not cast doubt on the claim that introspection reveals the essences of specific phenomenal properties. So, it seems that the thesis of revelation is not only intuitive, but also defensible and compelling.

Indeed, the fact that the intuitiveness of revelation is not easily explained in a way that is compatible with the physicalist ontology, as we saw in chapter 7, and the fact

that the thesis of revelation is not vulnerable to serious objections, as we saw in chapter 8, also go some way towards explaining the persistence of our intuition of dualism. Were the thesis of revelation easily explained away or defeated, our intuition of dualism would perhaps not run so deep.

We are now faced with the large-scale question of which metaphysical theories of consciousness can accommodate the thesis of revelation. We have already seen that revelation is incompatible with standard versions of physicalism (chapter 5), and is also in tension with grounding physicalism and some versions of Russellian monism (sections 3 and 4 of chapter 8). While many theories are ruled out by revelation, property dualism, as we saw in section 4 of chapter 8, is not. One can consistently maintain that on the one hand, the essences of phenomenal properties are revealed by experiences with these properties (revelation), and that on the other hand, phenomenal properties are fundamental and not physical/functional (dualism).¹

Now, it might turn out that property dualism, which is compatible with revelation, faces serious objections and fares substantially worse than theories that are incompatible with revelation. In such a case, one might need to deny revelation despite its intuitiveness. The main problem faced by dualists is the problem of mental causation – how does the mental have causal efficacy in the physical world? In contemporary philosophy, the problem is often known as ‘the exclusion problem’ (Kim 1993a, 1993b,

¹ There is a question as to whether emergentism is a form of property dualism, and in particular whether phenomenal properties, according to emergentism, count as *fundamental* if they are emergent from physical properties with metaphysical necessity. Here I shall count emergentism as a form of dualism. On emergentism, though phenomenal properties only occur when a physical system reaches a certain complexity, they are ungrounded and are thus fundamental in that sense. More on this in section 2.

1998, 2005). It questions the idea that phenomenal properties, if they are distinct from physical properties, can have causal effects, given that all the causal work can in every case be done by physical properties alone. The topic of mental causation is vast and deserves dissertation-length treatment. Nevertheless, in the rest of this chapter, I shall draw on recent literature and make preliminary remarks towards the idea that the exclusion problem faced by dualists is not as bad as is usually supposed. As I shall argue, the so-called 'exclusion principle', which underlies the exclusion problem, lacks clear motivation. Given this conclusion, it is far from clear that property dualism can be quickly dismissed – we have preliminary reasons to think that it not only can do justice to the thesis of revelation but is also philosophically viable.

2. DUALISM AND THE EXCLUSION PROBLEM

Contemporary dualists are mainly property dualists rather than substance dualists and, for this reason, it is property dualism that will be the focus here. In the recent literature, property dualism has typically been conceived along the lines proposed by Chalmers in *The Conscious Mind*. According to this conception, phenomenal properties are related to physical/functional properties by *fundamental* psycho-physical laws which are merely *contingent* – they hold in some possible worlds but not others. These laws link two fundamentally distinct kinds of properties. Consciousness only co-exists with these physical/functional features in worlds like ours where such laws hold. This kind of property dualism, as we already saw in section 4 of chapter 8, is compatible with revelation.

While Chalmers' dualism has become the standard conception of property dualism, we must note that property dualism need not be conceived in that way. Chalmers' version of property dualism takes phenomenal properties to supervene on physical properties with nomological necessity. One can alternatively maintain dualism while taking the modal strength of the relevant supervenience to be metaphysical rather than nomological. In the literature, this version of dualism is sometimes known as 'radical emergentism' or simply 'emergentism'.² Like nonreductive physicalism, emergentism takes phenomenal properties to be distinct from, but to metaphysically supervene on, physical properties. Though similar to nonreductive physicalism, emergentism can clearly be distinguished from the latter (see Crane 2001, 2010; Jackson 2006; Stoljar 2008). If physicalism is formulated in terms of a grounding thesis where 'grounding' is understood as a non-causal explanatory notion, then nonreductive physicalism, as a version of physicalism, would be committed to the idea that phenomenal properties are grounded in physical properties – the fact that we are subjects that instantiate phenomenal properties is *in virtue of* or *because of* (in the non-causal sense) the fact that we are subjects that instantiate certain physical properties. Emergentism, as a version of dualism, would reject this grounding thesis – even if

² It is 'radical' because the emergence of the phenomenal from the physical is brute. One might distinguish this radical kind of emergentism from a weaker kind, according to which the phenomenal nomologically supervenes on the physical (van Cleve 1990; Wong 2010; McLaughlin 2008; Noordhof 2010). One might posit fundamental emergence laws or what British emergentist C. D. Broad (1925) called 'trans-ordinal laws', holding between physical properties on the bottom, and distinct emergent properties on the top. This version of emergentism about consciousness looks indistinguishable from Chalmers' property dualism. In this section, by 'emergentism', I mean the kind that subscribes to the metaphysical supervenience of the phenomenal on the physical.

phenomenal properties metaphysically supervene on physical properties, this does not mean that facts about our consciousness are grounded in physical facts.³ On emergentism, conscious facts seem to be brute and ungrounded.

Without adjudicating which version of dualism is preferable, let me now explain the exclusion problem faced by all versions of dualism. Let Q be a phenomenal property of a mental state and let S be a physical effect of Q. (Here we are assuming that the causal relata are properties.) Thus, we have the following intuitive claim about the causal efficacy of the phenomenal in the physical world:

- (1) Q causes S. (*Efficacy*)

Now, when one is in a mental state with Q, one is also in a particular brain state with physical property P,⁴ such that the instantiation of Q is necessitated or determined, either metaphysically or nomologically, by the instantiation of P. Given the principle of the causal closure of the physical, i.e. every physical effect has a physical cause, we have the following claim:

³ Here I have formulated the difference in terms of the metaphysical notion of 'grounding'. In the literature, there are broadly two approaches to elucidating the difference between emergentism and nonreductive physicalism. Crane (2001, 2010), for instance, adopts an epistemological approach. In chapter 2, we saw that physicalism claims not only that phenomenal properties metaphysically supervene on physical properties, but also that such metaphysical supervenience is reductively explainable. We saw that reductive explanations can take two different forms – by appealing to either conceptual analysis or a posteriori identities. On Crane's epistemological approach, emergentists, *contra* nonreductive physicalists, reject the claim that phenomenal properties can be reductively explained with reference to the basal, physical properties on both accounts of reductive explanation. Stoljar (2008) opts for a metaphysical approach to distinguish the two views. According to Stoljar, emergentists and nonreductive physicalists actually mean different things when they both claim that 'the mental is distinct from the physical'. By distinguishing different notions of 'distinctness', Stoljar delineates the difference between the two positions.

⁴ Here I use the term 'physical' to exclude the functional.

(2) P causes S. (*Closure*)

Dualists, as well as nonreductive physicalists, endorse the following:

(3) $Q \neq P$ (*Distinctness*)

Dualists hold (3) because Q is a phenomenal property and P is a physical property, and phenomenal and physical properties are fundamentally distinct for dualists. Nonreductive physicalists hold (3) because Q is a property that may be multiply realised and P is just one of its realisers. For instance, P could be the physical property of being an event of C-fibres firing, whereas Q could be a second-order property, i.e. the property of having some physical property or other that plays the pain role.

Now, the above three claims are in tension given what Kim calls 'the exclusion principle':

No single event can have more than one sufficient cause occurring at any given time – unless it is a genuine case of causal overdetermination. (Kim 2005: 42)

If the exclusion principle is framed in terms of properties, it says that two distinct properties cannot have the same effect, except in genuine cases of overdetermination. A classic example of genuine overdetermination is the case of an execution by firing squad in which a man is simultaneously shot dead by two separate bullets. Let's formulate the exclusion principle as the following:

(4) Nothing distinct from P causes S (except in cases of genuine overdetermination). (*Exclusion*)

Note that from claims (1), (2) and (3), we reach the conclusion that S has two distinct causes, Q and P. Given (4), both Q and P are causes for S only if it is a case of genuine overdetermination. But intuitively, cases of mental causation do not seem to be genuine cases of overdetermination, *a la* the firing squad. So we have (5):

(5) Q and P do not overdetermine S. (*No Overdetermination*)

The claims (1) to (5) are inconsistent. Given (4), the exclusion principle, one of the other four – Efficacy, Closure, Distinctness, or No Overdetermination – must be incorrect.

As already touched on in section 2 of chapter 2, the exclusion argument against dualism can alternatively be transformed into a causal argument for reductive physicalism (i.e. identity physicalism). Premises (1), (2), (4) and (5) feature as premises in the latter argument (see Papineau 2001; Stoljar 2010):

(1) Q causes S.

(2) P causes S.

(4) Nothing distinct from P causes S (except in cases of genuine overdetermination).

(5) Q and P do not overdetermine S.

(Conclusion) $Q = P$.

Regarding the exclusion problem itself, reductive physicalists, i.e. those who identify phenomenal properties with physical properties, reject (3). As a result, they do not face the exclusion problem. If phenomenal properties are physical properties then their efficacy is not excluded. Nonreductive physicalists, on the other hand, do face the

exclusion problem (Kim 1993b, 1998). In response, they typically reject (4), the exclusion principle itself (Yablo 1992; Shoemaker 2007; List and Menzies 2009; Wilson 2009, 2011). Dualists take various approaches. Many reject (1) and take the phenomenal to be epiphenomenal (Chalmers 1996; Jackson 1982). One could reject (2), and the principle behind it, which is the causal closure of the physical. One could also reject (5) by accepting the idea that mental and physical causes systematically overdetermine the physical effects in the same way that the two separate bullets overdetermine the death of the man executed by the firing squad. There is also the option to reject (4), the exclusion principle itself, and to argue that there can be two distinct sufficient causes for an effect even if it is not a case of genuine overdetermination.

So at a first pass, dualists at least seem to have various avenues open to respond to the exclusion argument.⁵ Here I shall set aside the first three approaches available to dualists, i.e. rejecting either (1) or (2) or (5), and focus exclusively on the fourth approach, i.e. denying (4), which confronts the exclusion problem head-on. If the exclusion principle is not even plausible, then dualists certainly aren't burdened with the exclusion problem. Indeed, one might conjecture that, if nonreductive physicalists can successfully resist the exclusion argument and defend the causal efficacy of the phenomenal by rejecting the exclusion principle, then it might be the case that dualists can do just the same.

⁵ Correspondingly, the causal argument for identity physicalism has many critics. See Tiehen (2018: 547) for a short survey on scepticism regarding the causal argument (see also Stoljar 2010).

Kim (2005: 51) takes the exclusion principle to be ‘virtually an analytic truth with not much content’. But it is far from clear that the principle is as obvious as Kim believes it to be. Arguing against the exclusion principle, some theorists appeal to specific conceptions of causation (e.g. counterfactual account, interventionism, etc.) (LePore and Loewer 1987; List and Menzies 2009); some appeal to specific conceptions of causal relations (e.g. as tropes, as events, etc.) (Robb 1997; Gibb 2004; Davidson 1970, 1993); some understand the relationship between distinct causes – mental and physical properties – in a way that does not exclude the mental from doing the causal work (e.g. Pereboom (2002) appeals to the idea that mental properties are *constituted* by physical properties). But one might think that all of these proposals are laden with heavy metaphysical assumptions which are controversial. In the following, I shall explore a different way of denying the exclusion principle, which makes minimal metaphysical assumptions. This way of rejecting the exclusion principle crucially relies on drawing distinctions with respect to the notion of ‘distinctness’ which features in the exclusion principle. I shall explore this approach by taking an in-depth look at two noteworthy papers on the topic.

2.1. ÁRNADÓTTIR AND CRANE

In their paper ‘There is no Exclusion Problem’, Steinvör Thöll Árnadóttir and Tim Crane (2013) set out to show that nonreductive physicalists and emergentists do not face the exclusion problem. They do so by levelling a two-stage argument against the exclusion principle. In the first stage, they argue against the idea that the exclusion principle, as suggested by Kim, is an analytic truth. They (2013: 257) argue that ‘far from being an

analytic truth, the exclusion principle is not even plausible on its face. It conflicts with our causal judgements even before any physicalist commitments enter the picture'. Given the conclusion of the first-stage argument, namely, it is a very substantive claim which is not at all obvious, the exclusion principle should be supported by arguments. The second-stage of Árnadóttir and Crane's argument then naturally follows. In the second stage, they contend that the principle lacks independent motivation. Overall, they (2013: 251-2) argue that those who accept supervenience of the mental on the physical, and causal closure of the physical, but wish to hold the distinctness of the mental and the physical, do not face the exclusion problem because the exclusion principle is false. In this subsection, we consider in detail Árnadóttir and Crane's two-stage argument.

In the first stage of their argument, Árnadóttir and Crane (2013: 257) distinguish between '*independently* sufficient causes' and '*distinct* sufficient causes'. The former form a subcategory of the latter because distinct sufficient causes can also be *dependent*. In cases of genuine overdetermination, e.g. the firing squad, they argue, the numerically distinct causes are *independently* sufficient for the effect. So independently sufficient causes occur in cases 'where there are two or more causes, each of which would have been sufficient to bring about the effect *in the absence of the other*' (Árnadóttir and Crane 2013: 254). In contrast, the mental and the physical, while numerically distinct, are not independently sufficient causes for the effect, according to nonreductive physicalists and dualists – the mental *depends* on the physical. As Árnadóttir and Crane remark (2013:

255), '[f]or holding everything else fixed, you could not delete one of them *from a given context* without thereby deleting the other' (italics added).

Let Q be the phenomenal property of the *painfulness* of an event e, and let P be the physical property of *being an event of C-fibres firing*. A nonreductive physicalist who takes Q to be a functional property realised by P in humans, would say that if P is deleted in a given context then so is Q, and *vice versa*. In other words, if a human being stops undergoing an event of *C-fibres firing*, then she or he also stops undergoing a *painful* event, and *vice versa*. A dualist who takes Q and P to co-exist in humans also takes Q to depend on P. This dependence is in virtue of a nomic or metaphysical law holding between Q and P. Given this law, if we delete P in a given context, we thereby delete Q, and *vice versa*. So, a dualist would also say that if a human being stops undergoing an event of *C-fibres firing*, then she or he stops undergoing a *painful* event, and *vice versa*.

Given the distinction between *distinct sufficient causes* and *independently sufficient causes*, we should distinguish two readings of (4):

(4-num) Nothing numerically distinct from P causes S (except in cases of genuine overdetermination).

(4-ind) Nothing numerically distinct from P and sufficient for S *independently* of P causes S (except in cases of genuine overdetermination).

Árnadóttir and Crane, as we shall see later, problematise the exclusion principle formulated as (4-num), but do not problematise the principle understood as (4-ind). They do not problematise (4-ind) because (4-ind) is truistic, given the definition of

genuine overdetermination in terms of numerically distinct and independently sufficient causes (2013: 257). Cases of genuine overdetermination, e.g. the firing squad case, are precisely cases where one could delete one of the sufficient causes without thereby deleting the other. So, (4-ind), to quote Árnadóttir and Crane (2013: 257) is ‘a principle in the vicinity of the exclusion principle that is a better candidate for being “virtually an analytic truth”.’ In contrast, (4-num) is quite different from (4-ind) and is far from being an analytic truth as Kim claims.

The exclusion principle, as Kim intended it and as Árnadóttir and Crane understand it, is phrased in terms of the notion of numerically distinct sufficient causes, i.e. (4-num). It says that there can be numerically distinct causes only if they are also independently sufficient causes (i.e. cases of genuine determination). Árnadóttir and Crane argue that (4-num) does not cohere with our ordinary causal judgements, in which we often cite numerically distinct things as sufficient causes for the same effect, even though they are not independently sufficient. For instance, in the case of a hammer’s head making an indentation in some clay, they write, ‘[t]here seems no tension in saying both that the hammer caused the indentation in the clay and that its head did’ (2013: 257-8). The two relevant sufficient causes are not independent. If the hammer had not made an indentation in the clay, its head would not have made an indentation either. Conversely, if the head of the hammer had not made an indentation in the clay, the hammer would not have made an indentation either.

Another counterexample, pointed out by Christian List and Daniel Stoljar (2017), is Stephen Yablo’s (1992) case of the pigeon which has been trained to peck at red cards.

It seems natural to say that the redness of the card causes the pigeon to peck. But in a particular instance where the card is, say, scarlet, it also seems natural to say that the card's being scarlet causes the pigeon to peck. So, while being scarlet and being red are numerically distinct properties, it seems natural to say that both are sufficient causes of the pigeon's pecking even though the two causes are not independent (List and Stoljar 2017: 101). But given the exclusion principle understood as (4-num), only one of them can be the cause. This seems to be a counterintuitive consequence.

Given these intuitive cases against the exclusion principle formulated as (4-num), we can conclude – with Árnadóttir and Crane – that the principle fails to be obviously true; it does not even seem to be plausible on its face.

Kim precisely thinks that in cases where there are two putative (numerically) distinct sufficient causes that are not independent, the dependent (or supervenient) cause cannot be causally sufficient, as he (1998: 53) writes, 'the causal status of the dependent event is threatened by the event on which it depends'. So, are there reasons to hold the exclusion principle, i.e. (4-num), even though it fails to be analytic and obviously true? To put it differently, the concern here is whether there are reasons to think that the causal power of a *dependent* property is excluded by the property it depends on.

The second stage of Árnadóttir and Crane's argument contends that the exclusion principle, formulated as (4-num), lacks independent motivation. They point out that in motivating the principle, Kim relies on the idea that in order for a dependent property to count as a sufficient cause in addition to the physical cause, the dependent

property must bring some *additional* causal powers to the scene (Kim 1998: 53-55). But this idea, argue Árnadóttir and Crane, simply begs the question against those who take the mental to be irreducible to the physical but nevertheless efficacious. Those who maintain the (numerical) distinctness of the mental and the physical, i.e. (3), precisely deny that in the aforementioned instance of causation, Q brings causal powers in addition to P. It is not clear why, just because it does not bring additional causal powers to the scene, Q cannot count as a sufficient cause for S.

The exclusion principle, understood as (4-num) here, doesn't seem to be intuitive once we tease out the relevant claim at issue (given the first stage of the argument) and it lacks independent motivation (established by the second stage of the argument). Árnadóttir and Crane therefore conclude that nonreductive physicalists and emergentists, who maintain the irreducibility of the mental to the physical (that is, the numerical distinctness of the mental and the physical), simply do not face the exclusion problem because there isn't one.

2.2. LIST AND STOLJAR

Árnadóttir and Crane's strategy of defence against the exclusion principle is crucially underpinned by the distinction they draw between sufficient causes that are merely *numerically distinct* and those that are, in addition, *independently* sufficient. They (2013: 251) take their argument to show that there simply isn't an exclusion argument 'for those who deny the identification of mental causes with physical causes, while accepting supervenience and closure'. Although in their paper Árnadóttir and Crane are not

explicit about the modal strength of the supervenience at issue and are mainly concerned with the alleged exclusion problem faced by nonreductive physicalism and emergentism, their argument also seems to apply to Chalmers-style property dualism, which takes the mental to supervene nomologically rather than metaphysically on the physical.⁶

In a recent paper, Christian List and Daniel Stoljar (2017), also attend to different notions of distinctness – specifically, numerical and modal distinctness, where modal distinctness is defined as follows:

Two properties are *modally distinct* if and only if it is possible for the first to be instantiated and not the second *and vice versa*.⁷ (List and Stoljar 2017: 98)

They first respond to the exclusion argument on behalf of a nonreductive physicalist. They then argue that a Chalmers-style property dualist can also respond to the exclusion problem. Their project is in the same general spirit as Árnadóttir and Crane's, in the sense that they also do not make substantive metaphysical assumptions and appeal only to our ordinary causal judgements to dissolve the exclusion problem. In this subsection,

⁶ Furthermore, they note the similarity between aspects of their position and that of Mellor (1995: 103-5) who is critical of physicalist arguments for the claim that mental causes must be identical to, or (metaphysically) supervene on, physical causes.

⁷ In the terminology of Stoljar (2008), this version of distinctness is called 'strong' modal distinctness; for 'weak' modal distinctness, the 'and' is replaced by 'or'.

I shall draw on List and Stoljar's argument and further contend that the exclusion problem against dualism isn't significantly problematic.

Recall the exclusion problem, according to which the following five claims are inconsistent:

- (1) Q causes S. (*Efficacy*)
- (2) P causes S. (*Closure*)
- (3) $Q \neq P$. (*Distinctness*)
- (4) Nothing distinct from P causes S (except in cases of genuine overdetermination).
- (5) Q and P do not overdetermine S. (*No Overdetermination*)

The five claims are only inconsistent if the same notion of distinctness figures in (3), expressed by the symbol ' \neq ', and in (4), or if the notion of distinctness in (3) is more demanding than the notion in (4) (see List and Stoljar, 2017: 99-100). Thus, for example, the five claims are inconsistent if numerical distinctness figures in both (3) and (4), or modal distinctness figures in both (3) and (4), or modal distinctness figures in (3) and numerical distinctness figures in (4). There is no evident inconsistency if numerical distinctness figures in (3) and the more demanding modal distinctness figures in (4).

List and Stoljar's response to the exclusion argument against nonreductive physicalism is straightforward. According to nonreductive physicalism, phenomenal properties metaphysically supervene on physical properties. So, phenomenal property Q is not modally distinct from physical property P, and the notion of distinctness in (3)

must be numerical distinctness. For the exclusion problem to arise for nonreductive physicalism, the notion of distinctness in (4) must be numerical distinctness also. But as we saw from Árnadóttir and Crane, (4) understood as (4-num) is implausible.

List and Stoljar then extend their response to the exclusion argument levelled against Chalmers-style property dualism. They first distinguish two kinds of modal distinctness:

Two properties are *metaphysically modally distinct* ('metaphysically distinct' for short) if and only if it is *metaphysically* possible for the first to be instantiated and not the second *and vice versa*.

Two properties are *nomologically modally distinct* ('nomologically distinct' for short) if and only if it is *nomologically* possible for the first to be instantiated and not the second *and vice versa*.

Nomological distinctness is more demanding than metaphysical distinctness (which, in turn, is more demanding than numerical distinctness). If A and B are nomologically distinct, then they are also metaphysically distinct, but not *vice versa*. Standard Chalmers-style (rather than emergentist) property dualism takes phenomenal property Q and physical property P to be metaphysically distinct but not nomologically distinct. So, the notion of distinctness in (3) is metaphysical distinctness. For there to be an exclusion problem for Chalmers-style dualists, i.e. for there to be an inconsistency in the five claims above, the notion of distinctness in (4) has to be either numerical distinctness (4-num) or metaphysical distinctness (4-met):

(4-num) Nothing numerically distinct from P causes S (except in cases of genuine overdetermination).

(4-met) Nothing metaphysically distinct from P causes S (except in cases of genuine overdetermination).

We have already seen that (4-num) is implausible; the question is then whether (4-met) is plausible.

To bring out the implausibility of (4-met), List and Stoljar invite us to consider two properties, F and F*, which are nomologically co-extensive but metaphysically distinct. In all possible worlds that share the same laws as the actual world, F and F* are co-extensive; but there are (at least) two remote possible worlds (in which the actual laws do not all obtain) in one of which, something instantiates F but not F*, while in the other, something instantiates F* but not F. We are to suppose that F is causally sufficient for some effect E. Given (4-met), F* is not a sufficient cause of E.⁸ But this, argue List and Stoljar, shows (4-met) to be too strong. They write (2017: 104):

since F and F* are nomologically co-extensive, a natural thing to say is that either both are causally sufficient or neither is; but it is implausible that neither is causally sufficient; hence both are, and (4-met) should be rejected.

The thought here is that in at least some cases, properties that are nomologically co-extensive, and reveal their metaphysical distinctness only in remote counter-nomic possible worlds, are causally sufficient properties for the same effects.

⁸ Strictly speaking, the consequence is that F* is not a sufficient cause of E *unless it is a case of genuine overdetermination*. Here we are assuming that it is not like the firing squad case.

As an intuitive counterexample to (4-met), List and Stoljar provide the following

case:

[S]uppose that a junior lecturer loses his job because the university rejects his application for tenure; hence the university's making this decision—a property of the university—causes the lecturer's job loss. But suppose that, as a robust empirical regularity – a 'social-scientific law' – the committee that votes on these matters always consists of the most successful professors: they are reliably chosen to serve on the committee. Then it is also true that the most successful professors' judgment—a property of the professors—causes the lecturer's job loss.

The judgement of the most successful professors and the decision of the university are nomologically connected, and this connection is grounded in the relevant 'social-scientific law'. Such a law does not exist in remote counter-nomic possible worlds. So, the two properties at issue – a property of the university and a property of the professors – while not nomologically distinct, are metaphysically modally distinct. Given (4-met), it cannot be the case that both properties are sufficient causes. But intuitively, we want to say that both properties are causally sufficient for the effect – or, at the very least, we do not want to be bound to deny that both properties are causally sufficient for the effect. So this case is a counterexample to (4-met).

List and Stoljar (2017: 105) stress that they 'need not take a stand on whether [the following] principle is true':

If F is causally sufficient for E, and F is nomologically co-extensive with F*, then F* is also causally sufficient for E.

They regard the important point as being this:

It is often accepted that the non-reductive physicalist may say that a physical property and a mental property that is supervenient on it are both causally sufficient for some effect E on the grounds of both standing in the right nomological or counterfactual relation to E. It is then hard to see why the dualist may not say the analogous thing, where the metaphysical modality associated with supervenience is replaced by a nomological one, with everything else remaining structurally the same, especially as we focus on worlds that have the same laws of nature as ours. (List and Stoljar 2017: 105)

The take home message from List and Stoljar's paper is that the exclusion principle construed as either a claim about numerically distinct properties (4-num) or a claim about metaphysically distinct properties (4-met) is implausible. Since a Chalmers-style dualist takes phenomenal properties to be either numerically or metaphysically distinct from physical properties, and since the respective exclusion principles are implausible, such a dualist *prima facie* does not face the exclusion problem.

3. CHAPTER CONCLUSION

By appealing to different notions of 'distinctness', we have seen (in subsections 2.1 and 2.2) that the exclusion problem does not put as much pressure on dualists as has been commonly thought. The main lesson has been that versions of the exclusion principle that could potentially problematise dualism are not really compelling to begin with. Consequently, insofar as the exclusion problem is posed as an objection against dualism, dualists might have a plausible answer.⁹ Similarly, insofar as the exclusion problem

⁹ Of course, one might think that what has been said here does not amount to a full defence of dualism against the exclusion problem. One might think that there is some notion of distinctness that could feature in the proposition (3) in such a way that (a) the relevant distinctness claim is one that dualists hold but nonreductive physicalists reject and (b) the

against dualism can alternatively be formulated as the causal argument for reductive/identity physicalism, the causal argument might not be as compelling as its proponents have supposed.

One might think, of course, that dualism fails to be as ontologically simple and elegant as physicalism. Certainly, I have said nothing here in favour of the claim that dualism fares better than physicalism in the overall cost-benefit analysis; and what has been said here does not constitute a full defence of dualism. Nevertheless, we have preliminary reasons to think that dualism holds promise as a metaphysical theory of consciousness which can do justice to the thesis of revelation. In turn, there is no reason to reject revelation on the grounds that it leads to an unpromising metaphysical theory of consciousness. Indeed, philosophers who work on the metaphysics of consciousness should regard the thesis of revelation as a serious contender for truth.

corresponding exclusion principle (using the relevant notion of distinctness) is in fact plausible. Whether there is such a version of the exclusion principle that problematises dualism awaits further discussion.

BIBLIOGRAPHY

- Adams, R. M. (1995) 'Flavors, Colors, and God', in his *The Virtue of Faith and Other Essays in Philosophical Theology*, Oxford: OUP, pp. 243-62.
- Allen, K. (2011) 'Revelation and the Nature of Colour', *Dialectica* 65 (2): 153-76.
- — —. (2017) *A Naive Realist Theory of Colour*, Oxford: OUP.
- Alter, J. (1992) *The Wrestler's Body: Identity and Ideology in North India*, Berkeley: University of California Press.
- Alter, T. and Nagasawa, Y. (2012) 'What is Russellian Monism', *Journal of Consciousness Studies* 19 (9-10): 67-95(29).
- — —. (2015) *Consciousness in the Physical World: Prospects on Russellian Monism*, Oxford: OUP.
- Alvarez, M. (2010) *Kinds of Reasons: An Essay in the Philosophy of Action*, Oxford: OUP.
- Aristotle (1994) *Metaphysics. Book VII and VIII*, translated with a commentary by D. Bostock, Oxford: Clarendon Press.
- Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge and Kegan Paul.
- Árnadóttir, S. T. and Crane, T. (2013) 'There is no Exclusion Problem', in S. C. Gibb, E. J. Lowe, and R. D. Ingthorsson (eds.), *Mental Causation and Ontology*, Oxford: OUP, pp. 248-66.
- Balog, K. (2012) 'Acquaintance and the Mind-Body Problem', in S. Gozzano and C. S. Hill (eds.), *New Perspectives on Type Identity: The Mental and the Physical*, Cambridge: Cambridge University Press, pp. 16-42.
- Bayne, T. and Montague, M. (2011) 'Cognitive Phenomenology: An Introduction', in T. Bayne and M. Montague (eds.), *Cognitive Phenomenology*, Oxford, OUP, pp. 1-34.
- Bayne, T. and Spener, M. (2010) 'Introspective Humility', *Philosophical Studies* 20 (1): 1-22.
- Bealer, G. (1996) 'On the Possibility of Philosophical Knowledge', *Philosophical Perspectives* 10: 1-34.
- Bering, J. M. (2006) 'The Folk Psychology of Soul', *Behavioural and Brain Sciences* 29: 453-62.
- Block, N. (1978) 'Troubles with Functionalism', *Minnesota Studies in the Philosophy of Science* 9: 261-325.
- — —. (1995) 'On a Confusion about a Function of Consciousness', *Behavioral and Brain Sciences* 18: 227-47.

- Block, N. and Fodor, J. (1972) 'What Psychological States are Not', *The Philosophical Review* 81: 159-81.
- Block, N. and Stalnaker, R. (1999) 'Conceptual Analysis, Dualism, and the Explanatory Gap', *The Philosophical Review* 108: 1-46.
- Bloom, P. (2004) *Descartes' Babies*, London: Arrow Books.
- — —. (2007) 'Religion is Natural', *Developmental Science* 10 (1): 147-51.
- — —.. (2010) *How Pleasure Works: The New Science of Why We Like What We Like*, New York: Norton.
- Braddon-Mitchell, D. (2007) 'Against Ontologically Emergent Consciousness', in B. McLaughlin and J. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*, Oxford: Blackwell, pp. 287-99.
- Braddon-Mitchell, D. and Jackson, F. (2007) *The Philosophy of Mind and Cognition*, Oxford: Blackwell.
- Broackes, J. (1986) 'Belief *De Re* and *De Dicto*', *The Philosophical Quarterly* 36 (144): 374-83.
- Broad, C. D. (1925) *The Mind and its Place in Nature*, London: Kegan Paul.
- Byrne, A. (2006) 'Colour and the Mind-Body Problem', *Dialectica* 60 (3): 223-44.
- Byrne, A. and Hilbert, D. (2007) 'Colour Primitivism', *Erkenntnis* 66: 73-105.
- Camp, E. (2006) 'Metaphor and That Certain "Je Ne Sais Quoi"', *Philosophical Studies* 129: 1-25.
- Campbell, J. (1993) 'A Simple View of Colour', in J. J. Haldane and C. Wright (eds.), *Reality, Representation and Projection*, Oxford: OUP, pp. 257-68.
- — —. (2005) 'Transparency vs. Revelation in Color Perception', *Philosophical Topics* 33 (1): 105-15.
- Chalmers, D. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: OUP.
- — —. (2003) 'The Content and Epistemology of Phenomenal Belief', in Q. Smith and A. Jokic (eds.), *Consciousness: New Philosophical Perspectives*, Oxford: OUP, pp. 220-72.
- — —. (2010a) 'The Representational Character of Experience', in his *The Character of Consciousness*, Oxford: OUP, pp. 339-71.
- — —. (2010b) 'The Content of Phenomenal Concepts', in his *The Character of Consciousness*, Oxford: OUP, pp. 251-76.
- — —. (2014) 'Intuitions in Philosophy: A Minimal Defense', *Philosophical Studies* 171: 535-44.
- — —. (2016) 'The Combination Problem for Panpsychism', in G. Brüntrup and J. Jaskolla (eds.), *Panpsychism: Contemporary Perspectives*, Oxford: OUP, pp. 179-214.

- — —. (2018) 'The Meta-Problem of Consciousness', *Journal of Conscious Studies* 25 (9-10): 6-61.
- Chisholm, R. (1957) *Perceiving: A Philosophical Study*, Ithaca: Cornell University Press.
- Chudnoff, E. (2015) *Cognitive Phenomenology*, New York: Routledge.
- Cicero, M. T. (2006) *On Academic Scepticism*, trans. C. Brittain, Indianapolis: Hackett.
- Cohen, J. (2009) *The Red and the Real*, Oxford: OUP.
- Cohen S. M. (2016) 'Aristotle's Metaphysics', *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/win2016/entries/aristotle-metaphysics/>.
- Cottingham, J. (2000) 'Intentionality or Phenomenology? Descartes and the Objects of Thought', in T. Crane and S. Patterson (eds.), *History of the Mind-Body Problem*, London: Routledge, pp. 131-47.
- Crane, T. (2000) 'Introspection, Intentionality and the Transparency of Experience', *Philosophical Topics* 28: 49-67.
- — —. (2001) 'The Significance of Emergence', in G. Gillett and B. Loewer (eds.), *Physicalism and its Discontents*, Cambridge: Cambridge University Press.
- — —. (2005) 'Papineau on Phenomenal Concepts', *Philosophy and Phenomenological Research* 71 (1): 155-62.
- — —. (2010), 'Cosmic Hermeneutics vs. Emergence: The Challenge of the Explanatory Gap', in C. MacDonald and G. MacDonald (eds.), *Emergence in Mind*, Oxford: OUP, pp. 22-34.
- Crane, T. and Patterson, S. (2000) 'The Origins of Qualia', in T. Crane and S. Patterson (eds.), *History of the Mind-Body Problem*, London: Routledge, pp. 1-12.
- Damjanovic, N. (2012) 'Revelation and Physicalism', *Dialectica* 66 (1): 69-91.
- Dasgupta, S. (2014), 'The Possibility of Physicalism', *The Journal of Philosophy* 111 (9/10), 557-92.
- Davidson, D. (1969) 'The Individuation of Events', in N. Rescher (ed.), *Essays in Honor of Carl G. Hempel*, Dordrecht: Reidel, pp. 216-34.
- — —. (1970) 'Mental Events', in L. Foster and J. W. Swanson (eds.), *Essays on Actions and Events*, Oxford: Clarendon Press, pp. 207-24.
- — —. (1984) *Inquiries into Truth and Interpretation*, Oxford: OUP.
- — —. (1993) 'Thinking Causes', in J. Heil and A. Mele (eds.), *Mental Causation*, Oxford: Clarendon Press, pp. 3-18.
- Davies, M. (2000) 'Persons and Their Underpinnings', *Philosophical Explorations* 3 (1): 43-62.

- — —. (2015) 'Knowledge (Explicit, Implicit and Tacit): Philosophical Aspects', in J. D. Wright (ed.), *International Encyclopedia of Social and Behavioral Sciences*, Oxford: Elsevier Ltd.
- Davies, M. and Coltheart, M. (2000) 'Pathologies of Belief', *Mind and Language* 15 (1): 1-46.
- Davies, M. and Egan, A. (2013) 'Delusion: Cognitive Approaches – Bayesian Inference and Compartmentalization', in K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, and T. Thornton (eds.), *The Oxford Handbook of Philosophy and Psychiatry*, Oxford: OUP, pp. 688-727.
- Dennett, D. (1969) *Content and Consciousness*, London: Routledge.
- Descartes, R. (1985) *The Philosophical Writings of Descartes*, volumes 1-2, trans. J. Cottingham, R. Stoothoff, and D. Murdoch, Cambridge: Cambridge University Press.
- Díaz-León, E. (2014) 'Do A Posteriori Physicalists Get Our Phenomenal Concepts Wrong?', *Ratio* 27 (1): 547-62.
- Dietz, C. H. (2016) 'Are All Reasons Causes?', *Philosophical Studies* 173 (5): 1179-1190.
- Egan, A. (2008) 'Seeing and Believing: Perception, Belief Formation and the Divided Mind', *Philosophical Studies* 140: 47-63.
- Elpidorou, K. (2016) 'A Posteriori Physicalism and Introspection', *Pacific Philosophical Quarterly* 97: 474-500.
- Feigl, H. (1958) 'The "Mental" and the "Physical"', *Minnesota Studies in the Philosophy of Science* 2: 370-497.
- Fiala, B., Arico, A. and Nichols, S. (2011) 'On the Psychological Origins of Dualism: Dual-process Cognition and the Explanatory Gap', in E. Slingerland and M. Collard (eds.), *Creating Consilience: Issues and Case Studies in the Integration of the Sciences and Humanities*, Oxford: OUP, pp. 88-106.
- Fine, G. (2014) *The Possibility of Inquiry: Meno's Paradox from Socrates to Sextus*, Oxford: OUP.
- Fine, K. (1994) 'Essence and Modality', *Philosophical Perspectives* 8: 1-16.
- — —. (1995a) 'Ontological Dependence', *Proceedings of the Aristotelian Society* 95 (1): 269-90.
- — —. (1995b) 'Senses of Essence', in W. Sinnott-Armstrong, D. Raffman and N. Asher (eds.), *Modality, Morality and Belief: Essays in Honor of Ruth Barcan Marcus*, Cambridge: Cambridge University Press, pp. 53-73.
- Frankish, K. (2016) 'Illusionism as a Theory of Consciousness', *Journal of Conscious Studies* 23 (11-12): 11-39.
- Gelman, S. (2003) *The Essential child: Origins of Essentialism in Everyday Thought*, Oxford: OUP.

- Gert, J. (2008) 'What Colours Could Not Be: An Argument for Colour Primitivism', *The Journal of Philosophy* 105 (3): 128-55.
- Gertler, B. (2012) 'Renewed Acquaintance', in D. Smithies and D. Stoljar (eds.), *Introspection and Consciousness*, Oxford: OUP, pp. 89-123.
- — —. (2017) 'Self-Knowledge', *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/self-knowledge/>.
- Gewirth, A. (1943) 'Clearness and Distinctness in Descartes', *Philosophy* 18 (69): 17-36.
- Gibb, S. C. (2004) 'The Problem of Mental Causation and the Nature of Properties', *Australasian Journal of Philosophy* 82 (3): 464-76.
- Goff, P. (2011) 'A Posteriori Physicalists Get Our Phenomenal Concepts Wrong', *Australasian Journal of Philosophy* 89 (2): 191-209.
- — —. (2015) 'Real Acquaintance and Physicalism', in P. Coates and S. Coleman (eds.), *Phenomenal Qualities: Sense, Perception, and Consciousness*, Oxford: OUP, pp. 121-43.
- — —. (2017) *Consciousness and Fundamental Reality*, Oxford: OUP.
- — —. (forthcoming) 'Cosmopsychism, Micropsychism, and the Grounding Relation', in W. Seager (ed.), *The Routledge Handbook of Panpsychism*, Routledge.
- Golman, A. and Pust, J. (1998) 'Philosophical Theory and Intuitional Evidence', in M. DePaul and W. Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and its Role in Philosophical Enquiry*, Lanham, Maryland: Rowman and Littlefield Publishers.
- Gow, L. (2014) 'Colour', *Philosophy Compass* 9 (11): 803-13.
- Hausner, S. L. (2007) *Wandering with Sadhus: Ascetics in the Hindu Himalayas*, Bloomington: Indiana University Press.
- Hempel, C. (1965) *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press.
- Hill, C. (1991) *Sensations: A Defense of Type Materialism*, Cambridge: Cambridge University Press.
- — —. (2014) *Meaning, Mind, and Knowledge*, Oxford, OUP.
- Hodges, K. M. (2008) 'Descartes' Mistake: How Afterlife Beliefs Challenge the Assumption that Humans are Intuitive Cartesian Substance Dualists', *Journal of Cognition and Culture* 8: 387-415.
- Horgan, T. (1984) 'Functionalism and Token Physicalism', *Synthese* 59: 321-38.
- — —. (1993) 'From Supervenience to Superdupervenience: Meeting the demands of a Material World', *Mind* 102: 555-86.

- — —. (2012) 'Introspection and Phenomenal Consciousness: Running the Gamut from Infallibility to Impotence', in D. Smithies and D. Stoljar (eds.), *Introspection and Consciousness*, Oxford: OUP, pp. 405-22.
- Horgan, T. and Tienson, J. (2001) 'Deconstructing New Wave Materialism', C. Gillett and B. Loewer (eds.), *Physicalism and its Discontents*, Cambridge: Cambridge University Press, pp. 307-18.
- — —. (2002) 'The Intentionality of Phenomenology and the Phenomenology of Intentionality', in D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*, Oxford: OUP; pp. 520-33.
- Hornsby, J. (2000) 'Personal and Sub-personal; A Defense of Dennett's Early Distinction', *Philosophical Explorations* 3 (1): 6-24.
- Hume, D. (1739-40) *A Treatise of Human Nature*, ed. D. Norton and M. Norton (2000), Oxford and New York: OUP.
- Jackson, F. (1976) 'The Existence of Mental Objects', *American Philosophical Quarterly* 13: 33-40.
- — —. (1977) *Perception: A Representative Theory*, Cambridge: Cambridge University Press.
- — —. (1982) 'Epiphenomenal Qualia', *The Philosophical Quarterly* 32: 127-36.
- — —. (1986) 'What Mary Didn't Know', *The Journal of Philosophy* 83: 291-95.
- — —. (1998) *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford: OUP.
- — —. (2004) 'Mind and Illusion', in P. Ludlow, Y. Nagasawa, and D. Stoljar (eds.), *There is Something about Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, Cambridge, MA: MIT Press, pp. 421-42.
- — —. (2006) 'On Ensuring that Physicalism Is Not A Dual Attribute Theory in Sheep's Clothing', *Philosophical Studies* 131: 227-49.
- Jackson, M. B. (2018) 'Intuitions as Inferential Judgments', *Philosophical Issues* 28 (1): 7-29.
- Johnston, M. (1992) 'How to Speak of the Colors', *Philosophical Studies* 68 (3): 221-63.
- Kammerer, F. (2018a) 'Is the Antipathetic Fallacy Responsible for the Intuition that Consciousness is Distinct from the Physical?', *Croatian Journal of Philosophy* 18 (1): 59-73.
- — —. (2018b) 'Can You Believe it? Illusionism and the Illusion Meta-Problem', *Philosophical Psychology* 31 (1): 44-67.
- — —. (forthcoming) 'The Illusion of Conscious Experience', *Synthese*, DOI 10.1007/s11229-018-02071-y.
- Keil, F. C. (1989) *Concepts, Kinds, and Cognitive Development*, Cambridge, MA: MIT Press.
- Kim, J. (1966) 'On the Psycho-Physical Identity Theory', *American Philosophical Quarterly* 3 (3): 227-35.

- — —. (1993a) 'Can Supervenience and "Non-Strict Laws" Save Anomalous Monism?', in J. Heil and A. Mele (eds.), *Mental Causation*, Oxford: Clarendon Press, pp. 19-26.
- — —. (1993b) 'The Nonreductivist's Troubles with Mental Causation', In J. Heil and A. R. Mele (eds.), *Mental Causation*, Oxford: OUP.
- — —. (1998) *Mind in the Physical World: An Essay on the Mind-Body Problem and Mental Causation*, Cambridge: MIT Press.
- — —. (2005) *Physicalism, or Something Near Enough*, New Jersey: Princeton University Press.
- — —. (2008) 'Making Sense of Emergence', In M. Bedau (ed.), *Emergence: Contemporary Readings in Philosophy and Science*, Cambridge: MIT Press, pp. 127-54.
- Kriegel, U. (2013) 'A Hesitant Defense of Introspection', *Philosophical Studies* 165: 1165-1176.
- Kripke, S. (1980) *Naming and Necessity*, Cambridge: Harvard University Press.
- Lakoff, G. (1987) *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, Chicago: The University of Chicago Press.
- LePore, E. and Loewer, B. (1987) 'Mind Matters', *The Journal of Philosophy* 84: 630-42.
- Levine, J. (1983) 'Materialism and Qualia: The Explanatory Gap', *Pacific Philosophical Quarterly* 64: 354-61.
- — —. (1995) 'Qualia: Intrinsic, Relational or What?', in T. Metzinger (ed.), *Conscious Experience*, Paderborn: Schoenigh, pp. 277-92.
- — —. (2001) *Purple Haze: The Puzzle of Consciousness*, Oxford: OUP.
- — —. (2007) 'Phenomenal Concepts and Materialist Constraints', in T. Alter and S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford: OUP, pp. 145-66.
- Lewis, D. (1966) 'An Argument for the Identity Theory', *The Journal of Philosophy* 63 (1): 17-25.
- — —. (1972) 'Psychological and Theoretical Identifications', *Australasian Journal of Philosophy* 50 (3): 249-58.
- — —. (1980) 'Mad pain and Martian pain', in N. Block (ed), *Readings in the Philosophy of Psychology*, Cambridge, MA: Harvard University Press, pp. 216-22.
- — —. (1982) 'Logic for Equivocators', *Noûs* 16 (3): 431-41.
- — —. (1994) 'Reduction of Mind', in S. Guttenplan (ed.), *Companion to the Philosophy of Mind*, Oxford: Blackwell, pp. 412-31.
- — —. (1995) 'Should a Materialist Believe in Qualia?', *Australasian Journal of Philosophy* 73 (1): 140-44.
- — —. (1997) 'Naming the Colours', *Australasian Journal of Philosophy* 75 (3): 325-42.

- List, C. and Menzies, P. (2009) 'Nonreductive Physicalism and the Limits of the Exclusion Principle', *The Journal of Philosophy* 106: 475-502.
- List, C. and Stoljar, D. (2017) 'Does the Exclusion Argument Put Any Pressure on Dualism?', *Australasian Journal of Philosophy* 95 (1): 96-108.
- Loar, B. (1990/97) 'Phenomenal States', *Philosophical Perspectives*, 4: 81-108; with revisions in N. Block, O. Flanagan, and G. Güzeldere (eds.), *The Nature of Consciousness*, Massachusetts: MIT Press, pp. 597-616.
- Locke, J. (1689/1996) *An Essay Concerning Human Understanding*, abridged and edited by K. P. Winkler, Indianapolis/Cambridge: Hackett Publishing Company.
- Logue, H. (2016) 'Metaphysics of Color 2: Non-Physicalist Theories of Color', *Philosophy Compass* 11 (4): 220-31.
- Majeed, R. (2017) 'Ramsey Humility: the Response from Revelation and Panpsychism', *Canadian Journal of Philosophy* 47 (1): 75-96.
- Malcolm, N. (1964) 'Scientific Materialism and the Identity Theory', *Dialogue* 3: 115-25.
- Marmodoro, A. (2009) 'Do powers need powers to make them powerful? From Pandispositionalism to Aristotle', *History of Philosophy Quarterly* 26 (4): 337-52.
- Martin, M. G. F. (2002) 'The Transparency of Experience', *Mind and Language* 17 (4): 376-425.
- Marx, K. (1970) *Critique of Hegel's 'Philosophy of Right'*, trans. A. Jolin and J. O'Malley, Cambridge: Cambridge University Press.
- Massin, O. (2017) 'Bad by Nature: an Axiological Theory of Pain', in J. Corns (2017), *The Routledge Handbook of Philosophy of Pain*, New York: Routledge, pp. 321-33.
- McGinn, C. (1983) *The Subjective View*, Oxford: Clarendon Press.
- McLaughlin, B. (2001) 'In Defense of New Wave Materialism', in C. Gillett and B. Loewer (eds.), *Physicalism and its Discontents*, Cambridge: Cambridge University Press, pp. 319-30.
- — —. (2003) 'Color, Consciousness, and Color Consciousness', in Q. Smith and A. Jokic (eds.), *Consciousness: New Philosophical Perspectives*, Oxford: OUP, pp. 97-154.
- — —. (2008) 'Emergence and Supervenience', in Bedau and Humpreys (eds.), *Emergence: Contemporary Readings in Philosophy and Science*, Cambridge, MA: The MIT Press, pp. 81-98.
- Medin D. L. and Ortony, A. (1989) 'Psychological Essentialism', in S. Vosniadou and A. Ortony (eds.), *Similarity and Analogical Reasoning*, Cambridge: Cambridge University Press, pp. 179-95.
- Melnyk, A. (unpublished) 'The Sense of Incredibility: A Physicalist Explanation'.
- Montague, M. (2016) *The Given: Experience and its Content*, Oxford: OUP.

- Moran, R. (2001) *Authority and Estrangement: An Essay on Self-Knowledge*, New Jersey: Princeton University Press.
- Morris, K. (2018) 'What's Wrong with Brute Supervenience? A Defense of Horgan on Physicalism and Superdepervenience', *Analytic Philosophy* 59 (2): 256-80.
- Nagasawa Y. and Wager, N. (2016) 'Panpsychism and Priority Cosmopsychism', in G. Brüntrup and J. Jaskolla (eds.), *Panpsychism: Contemporary Perspectives*, Oxford: OUP, pp. 113-29.
- Nagel, T. (1974) 'What Is It Like to Be a Bat?', *The Philosophical Review* 83 (4): 435-50.
- — —. (1986) *A View From Nowhere*, Oxford: OUP.
- Newman G. E. and Knobe, J. (forthcoming) 'The Essence of Essentialism', *Mind and Language*.
- Nida-Rümelin, M. (2007) 'Grasping Phenomenal Properties', in T. Alter and S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford: OUP, pp. 307-36.
- — —. (2018) 'The Experience Property Framework: A Misleading Paradigm', *Synthese* 195: 3361-3387.
- Noordhof, P. (2010) 'Emergent Causation and Property Causation', in C. MacDonald and G. MacDonald (eds.), *Emergence in Mind*, Oxford: OUP, pp. 69-99.
- Palazitou, E. (2012) 'The Neurobiology of Depression', *British Medical Bulletin* 101: 127-45.
- Pandya, M., Altinay, M., D. A. Malone Jr., and A. Anand (2012) 'Where in the Brain is Depression?', *Current Psychiatry Reports* 14 (6): 634-42.
- Papineau, D. (2001) 'The Rise of Physicalism', in C. Gillett and B. Loewer (eds.), *Physicalism and its Discontents*, Cambridge: Cambridge University Press, pp. 3-36.
- — —. (2002) *Thinking about Consciousness*, Oxford: OUP.
- — —. (2007) 'Phenomenal and Perceptual Concepts', in T. Alter and S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford: OUP, pp. 111-44.
- — —. (2008) 'Explanatory Gaps and Dualistic Intuitions', in L. Weiskrantz and M. Davies (eds.), *Frontiers of Consciousness*, Oxford: OUP, pp. 55-62.
- — —. (2011) 'What Exactly is the Explanatory Gap?', *Philosophia* 39: 5-19.
- — —. (forthcoming) 'The Problem of Consciousness', in U. Kriegel (ed.), *The Oxford Handbook of the Philosophy of Consciousness*.
- Patterson, S. (2008) 'Clear and Distinct Perception', in J. Broughton and J. Carrierio (eds.), *A Companion to Descartes*, Blackwell, pp. 216-34.

- Peacocke, C. (1984/1997) 'Colour Concepts and Colour Experience', in A. Byrne and D. Hilbert (eds.), *Readings on Color Vol 1: The Philosophy of Color*, Cambridge, MA: MIT Press, pp. 51-66.
- — —. (1989) 'No Resting Place: A Critical Notice of the View from Nowhere', *The Philosophical Review* 98 (1): 65-82.
- — —. (2008) *Truly Understood*, Oxford: OUP.
- Pereboom, D. (2002) 'Robust Nonreductive Materialism', *The Journal of Philosophy* 99 (10): 499-531.
- — —. (2009) 'Consciousness and Introspective Inaccuracy', in S. Newlands and L. M. Jorgensen (eds.), *Metaphysics and the Good: Themes from the Philosophy of Robert Merrihew Adams*, Oxford, OUP, pp. 156-87.
- — —. (2011) *Consciousness and the Prospects of Physicalism*, New York: OUP.
- — —. (2013) 'Replies to Daniel Stoljar, Robert Adams, and Lynne Baker', *Philosophy and Phenomenological Research* 86 (3): 753-64.
- Perry, J. (2001) *Knowledge, Possibility, and Consciousness*, Cambridge, MA: MIT Press.
- Place, U. T. (1956) 'Is Consciousness a Brain Process?', *British Journal of Psychology* 47: 44-50
- Plato (1997) *Complete Works*, edited by J. M. Cooper, Indianapolis: Hackett.
- Pust, J. (2000) *Intuitions as Evidence*, New York: Garland Publishing.
- — —. (2017) 'Intuition', *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2017/entries/intuition/>>.
- Putnam, H. (1960) 'Minds and Machines', in S. Hook (ed.), *Dimensions of Mind: A Symposium*, New York: New York University Press, pp. 57-80.
- — —. (1967) 'Psychological Predicates', in W. H. Capitan and D. D. Merrill (eds.), *Art, Mind, and Religion*, Pittsburgh: University of Pittsburgh Press, pp. 37-48.
- Qu, H. (2017) 'Hume on Mental Transparency', *Pacific Philosophical Quarterly* 98: 576-601.
- Reuter, K. (2011) 'Distinguishing the Appearance from the Reality of Pain', *Journal of Consciousness Studies* 18 (9-10): 94-109.
- Reuter, K. Dustin, P., and Sytsma, J. (2014) 'Hallucinating Pain', in J. Sytsma (ed.), *Advances in Experimental Philosophy of Mind*, London: Bloomsbury Academic, pp. 45-74.
- Reuter, K. and Sytsma J. (2018) 'Unfelt Pains', *Synthese*. doi: 10.1007/s11229-018-1770-3.
- Reuter, K., Sienhold M., and Sytsma, J. (2019) 'Putting Pain in its Proper Place', *Analysis* 79 (1): 72-82.
- Robb, D. (1997) 'The Properties of Mental Causation', *The Philosophical Quarterly* 47 (187): 178-94.
- Robinson, H. (1994) *Perception*, London: Routledge.

- — —. (2011) 'Dualism', *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), <<http://plato.stanford.edu/archives/win2012/entries/dualism/>>.
- — —. (2015) 'Phenomenal Qualities: What They Must Be, and What They Cannot Be', in P. Coates and S. Coleman (eds.), *Phenomenal Qualities: Sense, Perception, and Consciousness*, Oxford: OUP, pp. 103-20.
- Sayre-McCord, G. (2015) 'Moral Realism', *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/fall2017/entries/moral-realism/>>.
- Schroer, R. (2010) 'Where's the Beef? Phenomenal Concepts as Both Demonstrative and Substantial', *Australasian Journal of Philosophy* 88 (3): 505-22.
- Schwitzgebel, E. (2008) 'The Unreliability of Naïve Introspection', *The Philosophical Review* 117(2): 245-73.
- — —. (2011) *Perplexities of Consciousness*, Cambridge, MA: MIT Press.
- Searle, J. (1997) 'Reductionism and the Irreducibility of Consciousness', in O. J. Flanagan, N. Block, and G. Guzeldere (eds.), *The Nature of Consciousness*, Cambridge, MA: MIT Press, pp. 451-60.
- Segal, G. (2007) 'Poverty of Stimulus Arguments Concerning Language and Folk Psychology', in P. Carruthers, S. Laurence, and S. Stich, (eds.), *The Innate Mind: Foundations and the Future, Volume 3*, Oxford: OUP, pp. 90-108.
- Shoemaker, S. (1994) 'Self-Knowledge and Inner Sense', *Philosophy and Phenomenological Research* 54: 249-314.
- — —. (2003) 'Content, Character, and Color', *Philosophical Issues* 13: 253-78.
- — —. (2007) *Physical Realization*, Oxford: OUP.
- Slingerland, E. (2011) 'Mind-Body Dualism and Two Cultures', in E. Slingerland and M. Collard (eds.), *Creating Consilience: Integrating the Sciences and the Humanities*, Oxford: OUP, pp. 74-87.
- Slingerland, E. and Chudek, M. (2011) 'The Prevalence of Mind-Body Dualism in Early China', *Cognitive Science* (2011): 1-11.
- Smart, J. J. C. (1959) 'Sensations and Brain Processes', *The Philosophical Review* 68 (2): 141-56.
- Sosa, E. (1970) 'Propositional Attitudes *De Dicto* and *De Re*', *The Journal of Philosophy* 67 (1970): 883-96.
- Spener, M. (2011) 'Disagreement about Cognitive Phenomenology', in T. Bayne and M. Montague (eds.), *Cognitive Phenomenology*, Oxford, OUP, pp. 268-84.
- Stanley, J. and Williamson, T. (2001) 'Knowing How', *The Journal of Philosophy* 97: 411-44.
- Stoljar, D. (2001) 'Two Conceptions of the Physical', *Philosophy and Phenomenological Research* 62: 253-81.

- — —. (2006) *Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness*, Oxford: OUP.
- — —. (2008) 'Distinctions in Distinction', in J. Kallestrup and J. Hohwy (eds.), *Being Reduced: New Essays on Causation and Explanation in the Special Sciences*, Oxford: OUP, pp. 263-79.
- — —. (2009) 'The Argument from Revelation', in D. Braddon-Mitchell and R. Nola (eds.), *Conceptual Analysis and Philosophical Naturalism*, Cambridge: MIT Press, pp. 113-38.
- — —. (2010) *Physicalism*, London: Routledge.
- — —. (2013a) 'Qualitative Inaccuracy and Unconceived Alternatives', *Philosophy and Phenomenological Research* 86 (3): 745-52.
- — —. (2013b) 'Four Kinds of Russellian Monism', in U. Kriegel (ed.), *Current Controversies in Philosophy of Mind*, London: Routledge, pp. 17-39.
- — —. (2016) 'The Semantics of "What it's like" and the Nature of Consciousness', *Mind* 125 (500): 1161-1198.
- — —. (2018) 'Review of Philip Goff, "Consciousness and Fundamental Reality"', *Notre Dame Philosophical Reviews*, [Online] <https://ndpr.nd.edu/news/consciousness-and-fundamental-reality/>.
- Smithies, D. and Stoljar, D. (eds.), (2011) *Introspection and Consciousness*, Oxford: OUP.
- Strawson, G. (1989) 'Red and 'Red'', *Synthese* 78: 193-232.
- Sundström, P. (2007) 'Colour and Consciousness: Untying the Metaphysical Knot', *Philosophical Studies* 136 (2): 123-65.
- — —. (2008) 'Is the Mystery an Illusion? Papineau on the Problem of Consciousness', *Synthese* 163: 133-43.
- — —. (2011) 'Phenomenal Concepts', *Philosophy Compass* 6 (4): 267-81.
- — —. (2017) 'How Physicalists Can – and Cannot – Explain the Seeming "Absurdity" of Physicalism', *Philosophy and Phenomenological Research*, doi: 10.1111/phpr.12394.
- Taylor, H. (2017) 'Powerful Qualities, Phenomenal Concepts, and the New Challenge to Physicalism', *Australasian Journal of Philosophy*, DOI: 10.1080/00048402.2017.1321678
- Tiehen, J. (2018) 'Physicalism', *Analysis* 78 (3): 537-51.
- Trogon, K. (2017) 'Revelation and Physicalism', *Synthese* 194: 2345-2366.
- Tye, M. (1992) 'Visual Qualia and Visual Content', in T. Crane (ed.), *The Contents of Experience*, Cambridge: Cambridge University Press, pp. 158-76.
- Van Cleve, J. (1990) 'Mind – Dust or Magic? Panpsychism versus Emergence', *Philosophical Perspectives* 4: 215-26.

- Veillet, B. (2015) 'The Cognitive Significance of Phenomenal Knowledge', *Philosophical Studies* 72 (11): 2955-2974.
- Weinberg, J. (2007) 'How to Change Intuitions Empirically Without Risking Scepticism', *Midwest Studies in Philosophy* 31: 318-41.
- Wierzbicka, A. (1989) 'Soul and Mind: Linguistic Evidence for Ethnopsychology and Cultural History', *American Anthropologist* 91 (1): 41-58.
- — —. (2006a) 'On Folk Conceptions of Mind, Agency and Morality', *Journal of Cognition and Culture* 6: 165-79.
- — —. (2006b) *English: Meaning and Culture*, New York: OUP.
- — —. (2010) *Experience, Evidence, and Sense: The Hidden Cultural Legacy of English*, Oxford: OUP.
- Wilson, J. (2009) 'Determinateness, Realization, and Mental Causation', *Philosophical Studies* 145: 149-69.
- — —. (2011) 'Non-Reductive Realization and the Powers-Based Subset Strategy', *Monist* 94: 121-54.
- Wilson, R. (2003) 'Intentionality and Phenomenology', *Pacific Philosophical Quarterly* 84 (4): 413-31.
- Wittgenstein, L. (1980) *Remarks on the Philosophy of Psychology*, Volume 1, trans. G. E. M. Anscombe, Blackwell.
- Wong, H. Y. (2010) 'The Secret Lives of Emergents', in A. Corradini and T. O'Connor (eds.), *Emergence in Science and Philosophy*, New York: Routledge, pp. 7-24.
- Yablo, S. (1992) 'Mental Causation', *The Philosophical Review* 101 (2): 245-80.
- Yolton, J. (1977) *A Locke Reader*, Cambridge: Cambridge University Press.